Solving The Inverse Problem Of ECG

USING NUERAL NETWORK ASSUMING A SINGLE DIPOLE

SUPERVISED BY

Prof. ABDEL-RAHIM KHALIFA

Reconstructing of The Depolarization Pattern of The Heart From Body Surface Potentials Using Neural Network

Ahmed Essam Mohamed

Ahmed Attia Mortada

Ahmed Omar Ahmed

Asmaa Mohamed Mansour

Andrew Abd El messih Fakhry

Taha Magdy Abd El Moneim

Mahmoud Salah Ahmed

Yasmeen Hassan Hassan

Yasmeen Mohamed Mohamed

Supervisor: Prof. Dr. Abdel Rahim Khalifa

A Thesis Presented for the Degree of Bachelor of Electronics and Communication Engineering

Electronics and Communication Engineering Department

Faculty of Engineering

Alexandria University

Egypt

2021 - 2022

Contents

Figures

Chapter 1: Electrical System of The Heart

1.1. Introduction

The electrical system of the heart is critical to how it functions. It controls the electrical impulses that cause your heart to beat and their conduction, which organizes the beating of your heart.

This electrical conduction across the heart's pathway is what is visible when traced on an electrocardiogram (ECG). The ECG is also what allows irregularities in the heart's electrical system, and with them any related symptoms and medical conditions, to be assessed.

1.2. Depolarization and Repolarization

Figure 0-1Depolarization & Repolarization

At start, the heart cell is resting, meaning it' s not really doing anything, it is just waiting until it is told to do something. So, whenever we have a resting heart cell, it is negatively charged on the inside and that is referred to as polarized cell and the reason for this that this cell has different concentration of ions on the inside compared to the outside and the cell membrane around the cell isn' t permeable right now, it' s impermeable specifically to sodium ions. There is a lot of sodium ions outside the cell, and they would love to get on the inside.

So, whenever the cell is stimulated, it receives an electric jolt from the electric conduction system of the heart, the cell membrane is changed, and it be more permeable so the sodium ions now can get on the inside of the cell and change the state from being negative to positive. That process is termed ***depolarization.*** And whenever the cell becomes depolarized it contracts.

The cell doesn' t stay depolarized forever, it repolarizes then. The membrane becomes impermeable again and the sodium ions get out of the cell by the sodium potassium pump. So, the membrane voltage is regained again

Next step after contraction is relaxation. So, after we have contraction, the cell relaxes by repolarization.

1.3. Where Do Electrical Impulses Start in the Heart?

The heart generates its own electrical signal. This electrical signal is produced by a tiny structure known as the sinoatrial node, located in the upper portion of the right atrium. The right atrium is one of four heart chambers and related valves which include two atria at the top of the heart with two ventricles at the bottom.

From the sinoatrial node, the electrical signal spreads across the right atrium and the left atrium (the top two chambers of the heart), causing both atria to contract. This pushes their load of blood into the right and left ventricles, the bottom two chambers of the heart.

The electrical signal then passes through the AV node to the ventricles, where it causes the ventricles to contract in turn.

The cardiac electrical signal controls the heartbeat in two ways. First, since each electrical impulse generates one heartbeat, the number of electrical impulses determines the heart rate. In normal sinoatrial rhythm, that rate will be between 60 and 100 beats per minute.

The sinoatrial node signal also controls electrical conduction of the heart's steps as it "spreads" across the heart. It causes the cells of heart muscle to contract in the correct sequence and ensures regular, efficient, and coordinated heartbeats.

1.4. Components of the Heart's Electrical Signal

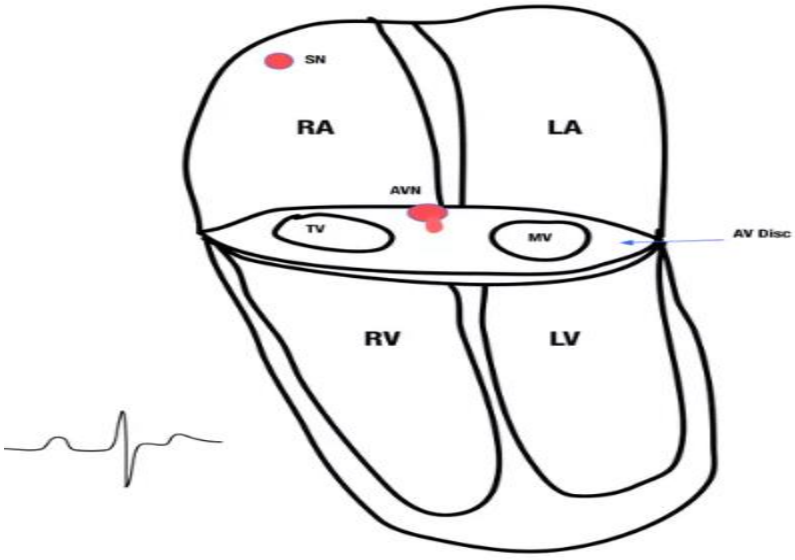

Figure 0-2 The components of the heart's electrical system

we can see the two atria and the two ventricles. Separating them is a layer of fibrous tissue, labeled AV disc. This tissue keeps the electrical signal passing through the AV node.

- SAN = sinoatrial node:
  It is a group of cells known as pacemaker cells, located in the wall of the right atrium of the heart. These cells can produce an electrical impulse (action potential) that travels through the electrical conduction system of the heart, causing it to contract.
- AVN = AV node:
  The atrioventricular node or AV node electrically connects the heart's atria and ventricles to coordinate beating in the top of the heart; it is part of the electrical conduction system of the heart. The AV node lies at the lower back section of the interatrial septum and conducts the normal electrical impulse from the atria to the ventricles. The AV node is quite compact (~1 x 3 x 5 mm).
- RA = right atrium
- LA = left atrium
- RV = right ventricle
- LV = left ventricle
- TV = tricuspid valve (the valve that separates the right atrium from the right ventricle)
- MV = mitral valve (the valve that separates the left atrium from the left ventricle)

1.5. The Heart's Electrical Signal Spreads Across the Atria

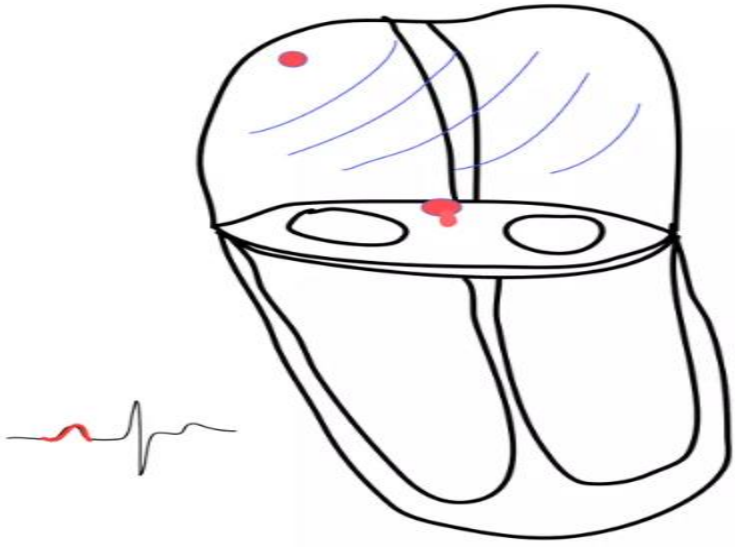

Figure 0-3 The electrical impulse originates in the sinoatrial node

The electrical impulse originates in the sinoatrial node. From there, it spreads across both atria (indicated by the blue lines in the picture), causing the atria to contract. This is referred to as atrial depolarization. As

the electrical impulse passes through the atria, it generates the so-called "P" wave on the ECG (The P wave is indicated by the solid red line on the ECG off to the left side).

There are some heart electrical system problems relating SA node:

- Sinoatrial bradycardia ("brady" means slow) is the most common cause of a low heart rate and is caused by the SA node firing at a reduced rate.
- Sinoatrial tachycardia ("tachy" means fast) refers to a rapid heart rate and can be caused by the SA node firing at an increased rate.

1.6. The Heart's Electrical Signal Reaches the AV Node

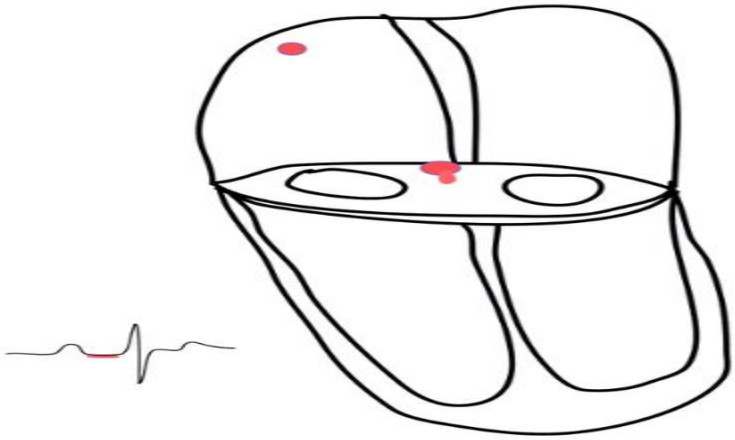

Figure 0-4 Heart's Electrical Signal Reaches the AV Node

When the wave of electricity reaches the AV disc, it is stopped, except in the AV node. The impulse travels through the AV node at a slow, controlled rate toward the ventricles. The solid red line on the ECG in this figure indicates the PR interval.

1.7. The Heart's Electrical Signal Passes to the Ventricles

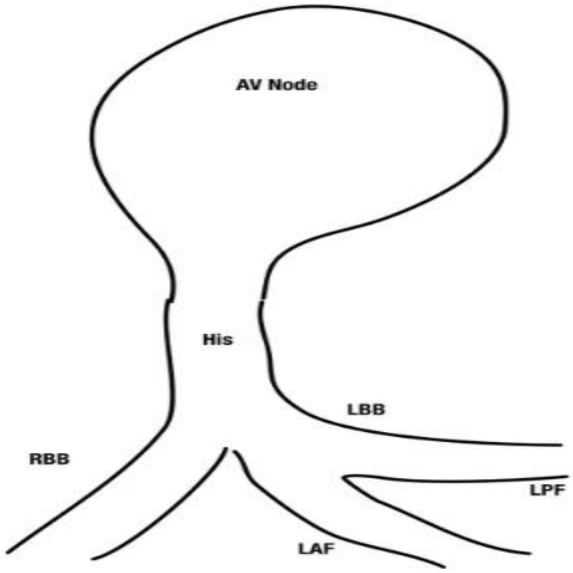

Figure 0-5 The specialized AV conduction system

The specialized AV conduction system consists of the AV node (AVN), the "His bundle," and the right and left bundle branches (RBB and LBB). The AV node conducts the electrical impulse to the His bundle. The His bundle passes the signal to the right and left bundle branches.

The right and left bundle branches, in turn, send the electrical impulse to the right and left ventricles, respectively.

Because the impulse travels only very slowly through the AV node, there is a pause in the electrical activity on the ECG, referred to as the PR interval. (The PR interval is illustrated on the ECG in Figure 1.6) This "pause" in the action allows the atria to contract fully, emptying blood into the ventricles before the ventricles begin to contract.

1.8. The Heart's Electrical Signal Spreads Across the Ventricles

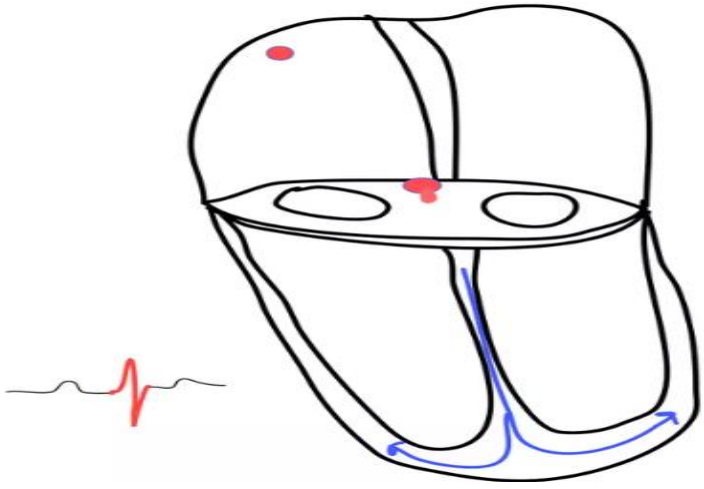

Figure 0-6 The electrical impulse spreading throughout the right and left ventricles

This figure shows the electrical impulse spreading throughout the right and left ventricles, causing these chambers to contract. As the electrical signal travels through the ventricles, it generates the "QRS complex" on the ECG. The QRS complex is indicated by the solid red line on the ECG to the left.

In this manner, the electrical system of the heart causes the heart muscle to contract and send blood to either the organs of the body (via the left ventricle) or to the lungs (via the right ventricle).

1.9. Normal ECG waves

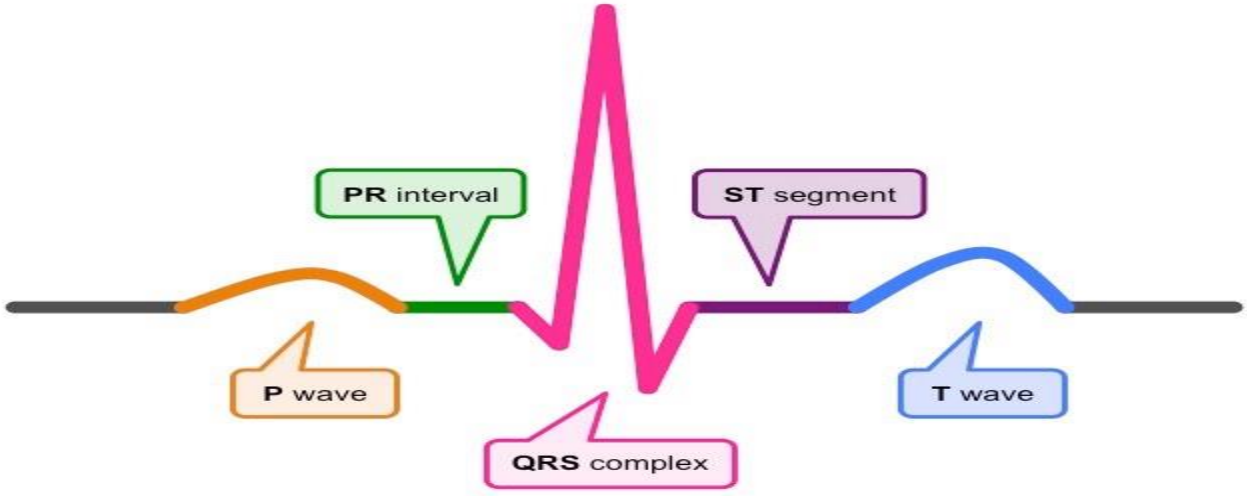

Figure 0-7 Normal ECG & Heart Components

As the heart undergoes depolarization and repolarization, the electrical currents that are generated spread not only within the heart, but also throughout the body. This electrical activity generated by the heart can be measured by an array of electrodes placed on the body surface. The recorded tracing is called an electrocardiogram (ECG). The different waves that comprise the ECG represent the sequence of depolarization and repolarization of the atria and ventricles.

1.9.1. P wave (Atrial Depolarization)

The P wave represents the wave of depolarization that spreads from the SA node throughout the atria and is usually 0.08 to 0.10 seconds (80-100 ms) in duration. The brief isoelectric (zero voltage) period after the P wave represents the time in which the impulse is traveling within the AV node (where the conduction velocity is greatly retarded) and the bundle of His. Atrial rate can be calculated by determining the time interval between P waves.

The period of time from the onset of the P wave to the beginning of the QRS complex is termed the PR interval, which normally ranges from 0.12 to 0.20 seconds in duration. This interval represents the time between the onset of atrial depolarization and the onset of ventricular depolarization. If the PR interval is >0.20 sec, there is an AV conduction block.

1.9.2. QRS complex (Ventricular Depolarization)

The QRS complex represents ventricular depolarization. Ventricular rate can be calculated by determining the time interval between QRS complexes. The duration of the QRS complex is normally 0.06 to 0.10 seconds. This relatively short duration indicates that ventricular depolarization normally occurs very rapidly. If the QRS complex is prolonged (> 0.10 sec), conduction is impaired within the ventricles.

1.9.3. ST segment

The isoelectric period (ST segment) following the QRS and ending at the beginning of the T wave is the time at which both ventricles are completely depolarized. Its duration is normally from 5 to 100 ms.

1.9.4. T wave (Ventricular Repolarization)

The T wave represents ventricular repolarization. Its duration is normally from 100 to 250 ms. Generally, the T wave exhibits a positive deflection. The reason for this is that the last cells to depolarize in the ventricles are the first to repolarize.

The T wave is longer in duration than the QRS complex that represents depolarization. The longer duration occurs because conduction of the repolarization wave is slower than the wave of depolarization. The reason for this is that the repolarization wave does not utilize the high-velocity bundle branch and purkinje system, and therefore primarily relies on cell-to-cell conduction.

1.9.5. QT interval

The QT interval represents the time for both ventricular depolarization and repolarization to occur, and therefore roughly estimates the duration of an average ventricular action potential. This interval can range from 0.20 to 0.40 seconds depending upon heart rate. At high heart rates, ventricular action potentials shorten in duration, which decreases the QT interval.

1.9.6. Atrial repolarization

There is no distinctly visible wave representing atrial repolarization in the ECG because it occurs during ventricular depolarization. Because the wave of atrial repolarization is relatively small in amplitude (i.e., has low voltage), it is masked by the much larger ventricular-generated QRS complex.

1.10. Conclusion

From the initiation of a heartbeat in the SA node, through contraction of the ventricles, the cardiac electrical system causes the heart to contract in a coordinated manner, maximizing the efficiency of the beating heart.

2. Chapter 2: The Forward Problem of Electrocardiogram

2.1. Introduction

In this chapter we describe a class of problems known collectively as the "forward problem of electrocardiography," which all share the goal of describing cardiac and torso electrical potentials starting from some description of electrical sources within the heart. To solve this forward problem, these electrical sources must be known beforehand, which may suggest a certain degree of artificiality, or at least impracticality, when viewed from the clinical context. The goal of clinical electrocardiography is to use the body-surface potentials from a patient to extract relevant parameters of the cardiac sources, which is the essence of the inverse problem of electrocardiography discussed in the following chapter. The forward problem, in contrast, has a more fundamental role in that it must capture the entire relationships between some description of the sources and the remote manifestations of cardiac bioelectricity. In its full scope, the forward problem begins with the membranes of the cardiac myocytes and goes to the body surface potentials; more limited formulations can start, for example, with potentials or activation times on the epicardial and endocardial surfaces, while others can start with descriptions of extracellular tissue potentials and predict epicardial electrograms. In all cases, a practical forward solution must first describe the sources, ideally in some way that strikes a compromise between spatial/temporal fidelity and tractability, i.e., it should be possible to measure or compute the values of interest. The forward solution must also capture in adequate detail the effect of the volume in which the sources are located, the "volume conductor," as its shape and electrical conductivity will determine the currents and potentials that form the solution of the forward problem. In this chapter, we present a rather broad view of the forward problem, one that includes all four components of a "versatile, present-day heart model" that Gulrajani outlined and are reflected in the goals of the Physiome and Cardiome projects. These components include the anatomical and physical substrate of the heart, the transmembrane action potential as the elemental bioelectric source of activation, the spread or propagation of this activation from cell to cell or element to element within the heart, and, finally, a volume conductor (typically the thorax) through which bioelectric currents pass from the heart to the outer surface, where they generate the electrocardiogram (ECG). One can examine each component of the complete problem with a range of approaches that include experiments with living tissues or mathematical and computational simulations, and complete coverage of all these options is well beyond the scope of this chapter. In this regard, our emphasis will be on mathematical and numerical approaches, although there will be some description of relevant findings from experiments. A notable challenge of a forward problem is its multiscale nature, i.e., a complete forward solution encompasses information from the scale of the ion-channel protein to the complete human thorax and from the nanosecond to minutes or even hours of time. One can measure or simulate transmembrane potentials from a cardiac cell (myocyte) or even unitary currents through single channels

of a cell membrane; however, both these approaches are intractable when the goal is to capture the behavior of the whole heart – there are thousands of ionic channels in a myocyte and billions of myocytes in a heart. Although mathematical models can include larger numbers of points to represent the heart than is possible with direct measurements, the density and complexity of representation are also limited by computer memory and computational capacity. Hence, one must accept simplifications and approximations that lead to workable formulations at the cost of detail and accuracy. Thus, for example, direct electrical measurements of the heart can come from extracellular potentials, sometimes only from the accessible surface(s), and at spacings in the range of millimeters to centimeters. At millimeter spatial resolution, the measurements can cover only small portions of the complete organ; to achieve more complete coverage results in a spatial resolution in the range of centimeters even with the most advanced measurement systems. Mathematical formulations of the forward problem are somewhat less restricted in the number of spatial locations that they can include but are also constrained in this case by computational resources. It is a general observation common to most biological system that the extreme complexity of the physiology of the heart at each scale requires simplification. It is not feasible to use the same detailed model of membrane behavior suitable for a simulation of a single cell when the goal is to study cardiac arrhythmias in a heart that contains billions of such cells. As a result, the choice of specific source representation and volume conductor resolution derive from the type of behavior one wishes to simulate and the questions such simulations might answer. In this chapter, we describe, for example, formulations that predict potentials on the body surface from those on the epicardial surface and that solve the resulting equations using a boundary element approach. There is simplification in that the problem includes only potentials on these two surfaces, which, in turn, has the advantage that we can represent each of these surfaces with many points (high spatial resolution) before exceeding the memory capacity of the computer. On the other hand, such a formulation forfeits direct information about cells or ionic channels or even the effects of the blood in the chambers of the heart. Although one pictures the forward solution as a physically or anatomically outwardly directed process, this is not necessarily the case. It is possible to formulate an inwardly directed forward problem in which the sources are in the heart tissues and the goal is to compute the electric potentials in the blood volume. This formulation is perhaps especially notable because it leads to a tractable inverse formulation that is the basis for a device that is the most widely used clinical application of the cardiac inverse problem [·]. The development of the forward problem has relied on both experimental and mathematical results and each approach has its respective strengths and limitations. The advantages of experimental approaches include the preservation, without simplification, of heart geometry and physiology, e.g., action potentials and spread of activation, and the ability to impose changes in this physiology through the use of drugs, artificial stimulation, temperature, mechanical load, or reduced coronary blood flow. Although experimental models contain the full complexity of the living tissues, measuring the parameters of interest is limited by physical access

and the maximum number – and thus spatial resolution and coverage – of simultaneous measurement channels. For example, it is not possible to capture a truly complete image of the time-dependent potentials or currents within a whole heart. Even measurements limited to, for example, the epicardial potentials can only occur invasively, thus disrupting the integrity of the physiologic volume conductor and the resulting body-surface ECG. Mathematical approaches remove the limitations of access and, to a certain extent, the number of parameters that one can monitor. The main challenge then becomes how to represent the true physiology in a realistic manner. Analytical approaches to mathematical solutions to the forward problem calculate the remote potentials from closed-form expressions for the cardiac sources. They offer great numerical precision, complete access at any desired resolution, and continuous variation of parameters. However, they are only possible under the most simplified geometric assumptions, e.g., that the heart and body are perfectly spherical or that lungs surround the heart in a concentrically spherical shape. They also implement an often highly simplified representation of the action potentials or spread of activation and sometimes compute body surface potentials not as time signals (ECG' s) but rather as sparse snapshots in time under specific conditions during which simple sources adequately capture cardiac fields. Numerical approximations of forward problems are certainly the most flexible and potentially powerful of all the options because they can, at least in principle, represent any sort of geometry in the form of discrete polygonal models and also any conceivable representation of bioelectric sources and spread of activation. Numerical forward problems in electrocardiography are generally unique in the sense that a specific set of source conditions leads to one and only one set of body-surface or epicardial potentials. As we describe in the following chapter, the same is not generally true of the inverse solution, e.g., multiple sets of cardiac source conditions can lead to the same set of body-surface potentials so that it may be impossible to determine which of the source conditions is correct. The presence of uniqueness should not, however, suggest that solving forward problems is trivial; there are considerable technical challenge and effort required to create the necessary geometric models, biophysical formulations, and numerical approximations. The geometry of the body is complex and contains regions of varying and even anisotropic electrical conductivity and there are many nonlinear relationships among relevant parameters. The complexity of simulations is also limited by computer capacity so that even relatively simplified simulations may take hours or even days to generate. Thus, the main areas of research in numerical forward problems involve the creation of realistic geometric models, the choice of appropriate simplifications of complex electrophysiology, and the search for more efficient means of implementing them. Although all three of these approaches – experimental, analytical, and numerical – have contributed to the knowledge of the forward problem, the dominant form in contemporary research is clearly the numerical simulation approaches with experimental results serving the essential role of validation.

2.2. Experimental Approaches

Here is a close link in the field of electrocardiography – and electrophysiology in general – between breakthroughs in measurement technology and new insights into the structure and mechanisms of the underlying behavior. This link exists at all levels of scale and has influenced the development of diverse forward problems in electrocardiography. The first ECG recordings were from Waller in the late nineteenth century but the real breakthroughs in ECG analysis and interpretation came, more than 20 years later, at least in part, because of Einthoven' s improvements to the string galvanometer that allowed rapid recordings from patients located remotely from the equipment. Near the other end of the size spectrum, the understanding of ion-channel function was first suggested by Hodgkin and Huxley based on their implementation of voltage clamp methods and then later expanded greatly because of the information that came from patch-clamp methods developed by Neher and Sakmann. The importance of such experimental techniques is further underscored by the fact that each of these breakthroughs resulted in a Nobel Prize for its innovators. From the enormous richness of experimental approaches and findings that have advanced electrocardiography and forward problems, we highlight here just a tiny sample of findings and information that is most relevant. Because of the large scope of the modern view of forward problems, such coverage must include cellular, tissue, whole heart, and thorax components. As we shall see, each of these fields of experimentation has a naturally synergistic counterpart in the domain of simulation and modeling and one cannot imagine contemporary research without close coupling between these domains.

2.2.1 Cellular Electrophysiology

The broad goals of electrophysiological measurements at the cellular level are to determine the resting and dynamic electric potentials across the cell membrane and to measure the associated currents that flow through the membrane. This current flow occurs through hundreds to thousands of ion channels, each of which belongs to one of tens of channel types. Each channel is composed of an opening or pore, surrounded by complex amino-acid helices that form several separate proteins of largely known composition. It is the characteristics of these proteins and the changes in structure that they undergo that ultimately give rise to ionic currents, changes in membrane voltage, and the driving forces for bioelectricity in the heart and thorax. These channel types differ in terms of their selectivity for particular ions, their electrical conductivity, and their time and voltage-dependent activity. The characterization of these features in normal and diseased channels forms the central theme of a great deal of experimental studies. Measurements of transmembrane potential generally occur by means of electrodes placed inside individual cells, much as one measure potential difference between two locations in a circuit. Because cells in the myocardium are relatively small (roughly brick shaped with 100µm length and 10–20 µm sides), electrodes are created from small pieces of glass tubing pulled under heat to a diameter of under 1 mm.

Measurements of individual ion currents typically occur by means of slightly larger glass microelectrodes that attach to the membrane and isolate small numbers and even single channels. The measurements in this case are of the currents that flow through the small patch of membrane under the electrode tip. For use in forward problems, the most important variation on this basic theme is the voltage-clamp configuration, which uses either two separate electrodes or one that switches between functions at a frequency high enough for it to carry out t the roles of both separate electrodes.

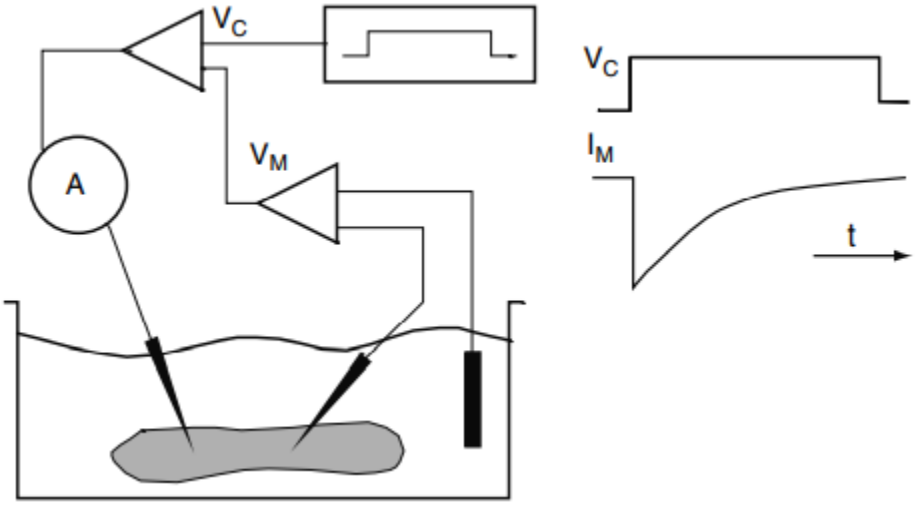

Figure 2-1 Schematic diagram of the voltage clamp procedure

Two electrodes impaled in the cell provide membrane potentials measurement and current injection. Comparison of the membrane voltage Vm and the command potential Vc creates a difference signal that injects measured current into the cell in order to maintain Vc = Vm. Time signals indicate the response of time- and voltage-dependent current to a step pulse of the control potential.)

As shown in > Fig. 2.1, the purpose of the voltage clamp paradigm is to hold, or clamp, the transmembrane voltage at a chosen value and then measure the ionic current required to maintain this potential. Because the voltage is held constant across the membrane, i.e., the voltage-dependent behaviors are fixed, the resulting current measurements reveal the temporal behavior of the channels. Another essential tool for studying cellular electrophysiology is a set of drugs that selectively block individual ion selective currents. Hodgkin and Huxley made use of TTX, a toxin derived from the puffer fish that very selectively blocks Na+ channels (and still causes many deaths each year among puffer-fish gourmands) and TEA, a blocker of some types of potassium channels. With selective channel blockers, it is possible to isolate individual currents and use voltage clamp to determine their unique voltage and time dependence. Thus, it was that Hodgkin and Huxley were able to determine the time- and voltage-dependent characteristics of sodium and potassium channels in the squid giant axon and both formulate and test their Nobel-prize winning

approaches to mathematically modeling the behavior of ion channels. The widespread availability of techniques from molecular biology and knowledge of the protein sequence of ion channels continues to shape the contemporary approach to cellular electrophysiology. It is now possible to alter in very controlled ways the sequence of amino acids that make up ion channels and to them embed these modified channels into selected cell types and then study their behavior. In this way, it is possible to address questions in basic cellular electrophysiology and also to create experimental models of a wide range of disease states. Such changes in structure are possible through direct cloning of ion channels, but also by manipulating the genetic code of (usually) mice to create viable transgenic species lines that exhibit specific disease states. As just one example of this powerful approach, researchers altered the structure of the sarcolemma ATP-sensitive potassium (KATP) channels and compared them with normal (wild type) channels in cells and intact transgenic mice. These studies showed that the KATP channel is primarily responsible for the changes in action potential morphology during myocardial ischemia that led to the shifts in body-surface ST-segment potentials that are the most common diagnostic feature for myocardial ischemia and infarction.

2.2.2 Tissue and Whole Organ Experiments

The science and technology of experimental approaches for myocardial tissue and the whole hearts is enormous and includes applications in both basic research and day-to-day clinical practice. Here we outline the general conceptual framework for such measurements as they relate to the description of bioelectric sources and to some specific approaches that are relevant to any discussion of the electrocardiographic forward problem.

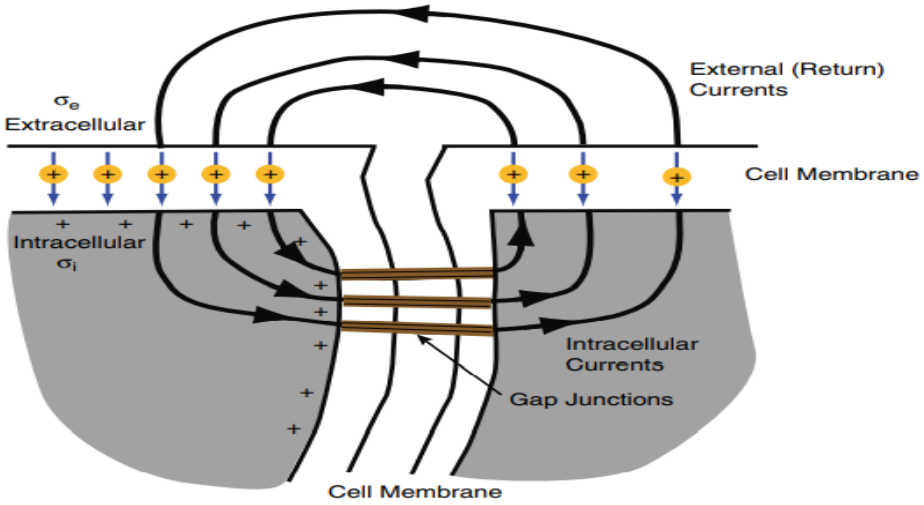

Figure 2-2 Schematic diagram of currents flowing between cells

The figure shows the interface between two cells with cell membrane and gap junctions linking the ends of cells. The left-hand cell is depolarized so that its intracellular space becomes positive with respect to the right-hand cell; hence, current flows directly from one intracellular space to the other through the gap junctions. Extracellular currents form the necessary return path for these currents. Conductivities σi and σ e characterize the intracellular and extracellular space, respectively.

The main goal of experimental approaches at the tissue and whole-organ level is to measure features of cell-to-cell coupling and observe behaviors that arise from the integration of many cells into the complex, three-dimensional structure of the heart. The measured quantity is almost invariably voltage coming from the extracellular current, i.e., the return current that flows between cells through the extracellular space. In fact, it is the qualitative appreciation of the difference between intracellular and extracellular measurements and currents that is perhaps most relevant to any discussion of the electrocardiographic forward problem. This conceptual framework provides the bridge between ionic currents and the macroscopic bioelectric fields that is central to any discussion of cardiac sources, which we describe in > Sect. 2.3. > Figure 2.2 shows schematically the relationship between ionic currents and the passive currents that flow between cells and into the extracellular space. In the figure, the left-hand cell is depolarized while the right-hand cell is at rest, thus producing the potential difference that is always necessary for current to flow from cell to cell, and by extension, from one part of the heart to another. Current is able to flow easily between neighboring cells because of the presence of gap junctions so that the downstream cell then also begins to accumulate charge. This initial charge accumulation, in turn, provides capacitive current that appears to flow (as all capacitive currents appear to flow – there is no ion flow across the membrane but rather a movement of ions on both sides that is the current) into the extracellular (interstitial) space. This extracellular, or return, current flows through the interstitium, which has a discrete conductivity, σe, and thus generates local voltage differences that one can measure.

2.2.2.1 Lead Systems and Electrode Arrays:

There are two basic approaches to measuring cardiac activity: unipolar and bipolar leads. Both have physical implementations in the form of contact electrodes. A lead is the potential-difference time-signal recorded between two locations in or around the heart or body; signals from the body surface are electrocardiograms (ECG' s) and signals from the heart are electrograms (ECG' s). Unipolar leads in measurements of cardiac bioelectricity represent the potential difference between one site on or near the heart and a reference, or indifferent electrode typically located remotely from the heart. The actual electrodes can be a small metal pellets, disks, or uninsulated lengths of wire, placed in direct contact with myocardium. The electrode is connected to the noninverting input of an electronic differential amplifier,

while the inverting (negative) inputs of all the amplifiers are connected to the indifferent electrode. Thus, the potential measured by each amplifier represents the difference of voltage between an individual electrode recording site and the reference site. The main advantages of unipolar leads are that they share a common reference and thus provide the necessary information for comparing values over different leads and hence over space. The recording of voltage over space is known as cardiac potential mapping. A disadvantage of unipolar leads is that the signals contain both local information, attenuated fields from remote events, and measurement noise, and one often wishes to separate local from remote events. The signals in Panels A and B in > Fig. 2.3 are examples of unipolar electrograms. Bipolar leads consist of two, closely spaced electrodes, with one connected to the inverting and the other to the noninverting input of an amplifier and have the advantage of increased sensitivity to local activity while reducing the effects from distant activity and noise. The bipolar signal represents the algebraic difference between the unipolar signals that would have been recorded from the separate electrodes. This subtraction removes common information or similarities in the two unipolar signals, leaving only those aspects of the unipolar signals that are different. Such differences, in turn, are the result of local events so that bipolar leads emphasize the passage of the activation wavefront and may improve detection of activation. One weakness of bipolar leads is that they do not share a common reference and thus cannot be compared with other leads in any sort of mapping based on amplitude or time course of the signals. Figure 2.3 shows two pairs of unipolar EGs recorded from closely spaced electrodes (Panels A and B) as well as their differences – the bipolar signals that would have been recorded from them (Panels C and D). from the perspective of the electrocardiographic forward problem, another salient features of lead type are that unipolar and bipolar leads sense different information, and each type provides data for a different formulation of the forward problem.

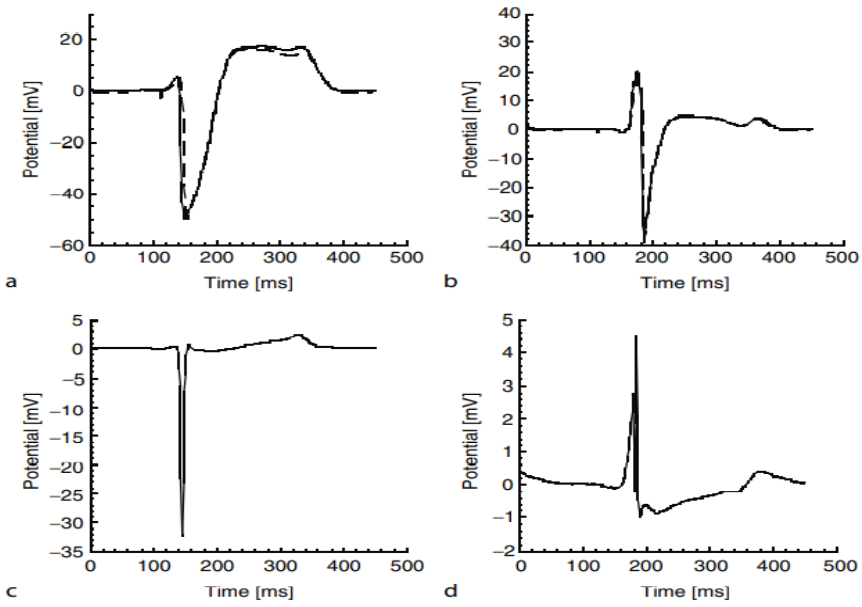

Figure 2-3 A sample of unipolar electrograms and the bipolar leads derived from their difference

(a and b) each contain two very similar unipolar electrogram from neighboring sites in a 1,200-lead epicardial array. (c and d) contain the bipolar leads derived from the two unipolar signals in (a and b) respectively. Note the different vertical scales in each plot.

Unipolar leads recorded from the epicardial surface(s) of the heart provide direct information for the form of the forward problem based on surface potentials sources. Bipolar leads, on the other hand, provide direct source information for forward models based on activation times. It is imperative to note that one can fairly easily derive activation times from unipolar electrograms, but that it is impossible to derive potential maps from bipolar leads. A fundamental question in any mapping and forward or inverse modeling application is the number and location of electrodes required. From a purely experimental perspective, the best number/location configuration of electrodes is based on a combination of features such as the desired biomedical goal, the accessibility of the heart to measurements, the types of electrodes available, and the capabilities of the acquisition system. The question of the minimum required spatial resolution has not been resolved completely, but current consensus indicates that a spacing of ≈ 2.5 mm is required to capture the details of cardiac activation. This resolution assumes simple linear interpolation between measurement sites, either through explicit interpolation or implicitly in the visualization or signal processing required to identify features of interest. There are, however, more sophisticated forms of interpolation that allow, for example, lower density measurements (≈ 10–15 mm) on the epicardial surface and are still capable of reconstructing activation wavefronts. In an even more extreme case, if a set of high-resolution training data is available, it is possible using statistical estimation techniques to reconstruct

activation maps on the entire ventricular surface from a very unevenly spaced set of only electrodes restricted to the coronary veins. A selection of typical electrode configurations of unipolar and bipolar lead configurations used in contemporary electrophysiology, a subset of which is shown schematically in > Fig. 2.4, includes the following:

1. Epicardial arrays sewn into nylon socks that cover some or all of the ventricles.
2. Rigid plaque arrays with regular electrode spacing that cover 1-10 cm2 areas of the ventricles or atria; 3. Sets of transmural plunge needles with 3-12 electrodes in each needle.
3. Catheters with 1-16 electrodes that are placed in the ventricles, atria, or coronary veins.
4. Basket catheters consisting of 4-6 strands, each containing electrodes for insertion into the right or left atrium.
5. 5. An inflatable, multielectrode balloon or large bore catheter with 10-80 electrodes that can be introduced into the ventricles through an incision or through the coronary vessels.

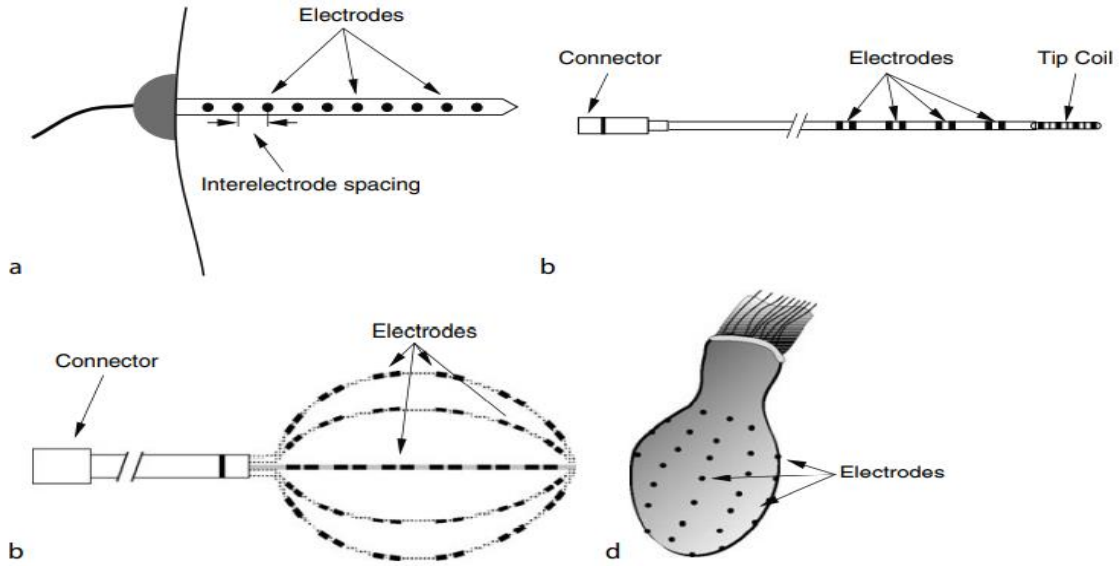

Figure 2-4 A sample of electrode arrays commonly used in cardiac mapping

(a) plunge needle with ten individual electrodes; (b) multielectrode catheter arranged in four bipolar pairs; (c) basket catheter with six wands, each containing six bipolar pairs; and (d) inflatable sock array for surgical insertion into the left ventricle.

2.2.3 Physical Model and Phantoms of the Thorax

Physical models of cardiac sources and volume conductors have existed almost as long as the field of electrocardiography and much of the original understanding of the electrocardiographic field came from studies with such models. One can argue whether the ultimate goal of such models has been to solve the forward problem or the inverse problem of electrocardiography; investigators simply set out to understand the relationship between some form of cardiac source and the resulting potentials and currents on and in the torso volume conductor. And here, we introduce some of them

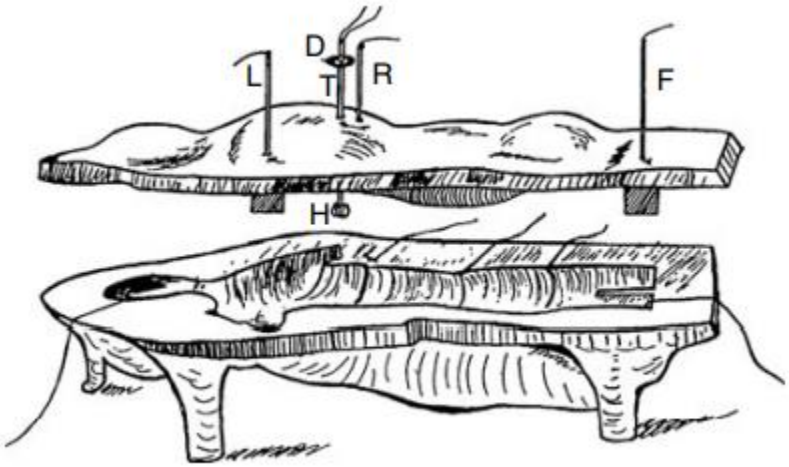

Figure 2-5 Electrolytic tank from Burger and van Milaan

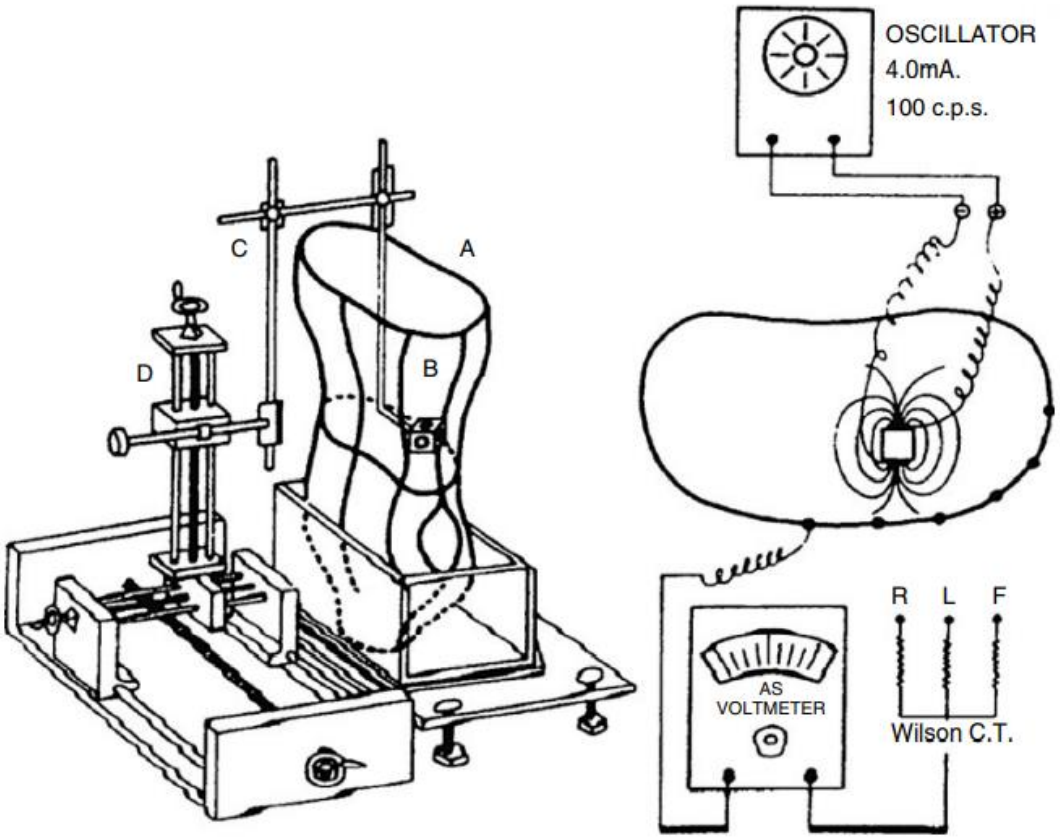

Figure 2-6 Electrolytic tank from Nagata

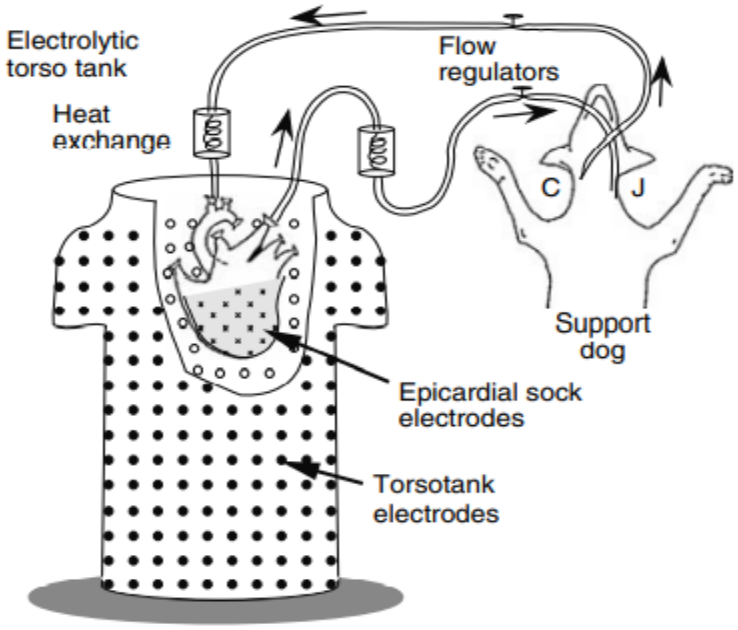

Figure 2-7 Physical model of forward problem originally devised by Taccardi et al

With an isolated, perfused dog heart suspended in the electrolytic tank. Recording electrodes consist of 192-384 tank surface electrodes and a 64-490 lead epicardial sock array.

2.3. Modeling Cardiac Bioelectricity

Creating and using simulations of cardiac bioelectricity has a long and rich history and also covers an enormous range of scale and mathematical sophistication. In this chapter, we provide only a brief overview of this material and focus on the methods that are most relevant in the contemporary research and that drive the forward problem in electrocardiography. As we have defined it, the scope of the forward problem is also very large, and therefore, we include at least some coverage over the complete spectrum from ionic channels to whole hearts. The unifying goal of simulating cardiac sources is to represent in some compact and mathematically defined way the currents and electric potentials that the heart produces. These representations may be based on explicit anatomical information and even require geometric models of cells, tissue, or the whole heart. Source models may, however, also bear little resemblance to realistic anatomy, but be abstractions of real biophysical sources that are useful only when the distances from the source are large compared with the extent of the source. One can view such a range of representations as similar to describing a source of light either in terms of the actual filament shape of the bulb or, simplified, as a point source radiating equally in all directions; both representations are useful at some scale but not usually across the full range of possible scales. Similarly, representations of cardiac bioelectricity even at the same scale may differ in degree of accuracy or sophistication and they may also vary in terms of the range of situations they can simulate. For example, a source model may create highly realistic potentials for a normal heartbeat but be completely inadequate for pathological states. In all cases, source representations are approximations and hence reflect some degree of compromise, usually in order to be mathematically or computationally tractable.

2.3.1. Models of Cardiac Myocytes

The requirement of a model of cardiac cells is that they simulate with acceptable fidelity both the resting and dynamic behavior of the cell. Resting potentials depend on the concentrations of ions on both sides of the cell membrane and on the resting conductivity of ion channels, and a complete model of the cell will reflect these dependencies and even allow for variations in these parameters that reflect pathophysiology, e.g., a cell model should allow the resting potential to rise (depolarize) with elevation in extracellular potassium ([K+ e]), as a real cell does during acute ischemia. Accurate representation of the action potential of the cell is a much more challenging requirement, especially given that, again, this behavior must reflect a wide range of variations in all the relevant ion channels and their time and voltage dependence. The action potential of a myocyte is the sum of tens of different types of channels arranged in densities that vary with the cell type, species, and even location in the same heart. Each ion channel, in turn, opens and closes in a stochastic manner and thus allows current to flow in a complex time, voltage, and ligand dependent manner, described ultimately by the set of proteins that make up the channel

structure. An ideal model of the cell would allow predictions of ionic currents and action potential shape on the basis of variation in the underlying protein structure, i.e., the sequence of amino acids that make up the proteins, determined by the DNA of the cell nucleus. Such a complete model does not yet exist but there do exist frameworks in which to approach the required (or desired) sophistication. Moreover, there are formalisms that are based on at least reasonable simplifications and ensemble averages of the individual ion channel characteristics that seek to predict the behavior of pieces of cell membrane and whole cells. Even more simplified models also exist that approximate the essential parameters of action potential behavior and are driven by empirical mathematical formulations. Here, we describe briefly these first category of cell models as they represent the far dominant form in forward solution formulations.

2.3.1.1. Biophysically Based Models

There are two levels of biophysically based modeling approaches in common use today: one that describes the opening and closing of individual ion channels and the other that computes currents and voltages for the entire cell. Both of them share the characteristic that they appeared before there was clear experimental proof of the behavior and especially the structure that they attempt to simulate. In this sense they illustrate one of the most powerful applications of simulations, that is to start with a concept of the underlying mechanism and then create a quantitative model that reflects this concept as a means of testing it against measured data. The Hodgkin–Huxley formalism was the result of breakthroughs in both measurement and theory that occurred in the period of rapid progress that followed World War II. Hodgkin and Huxley made use of the electronic circuitry and devices developed during the war to implement and then apply the voltage-clamp technique the squid giant axons . In order to describe first conceptually and then quantitatively the results of these experiments, they proposed the idea of ion channels, that is, openings in the axon membrane that were selective to specific ions and that had time and voltage dependencies that voltage clamp allowed them to investigate. Starting from first-order kinetic equations common in physical chemistry and ordinary differential equations frequently used to describe simple rate dependencies, they adjusted parameters in order to fit the measured data and presented calculations to support these concepts . Subsequent experiments, requiring numerous technological breakthroughs and performed by their successors, proved that the concept of ion channels was sound and led to a Nobel Prize for Hodgkin and Huxley. This simulation formalism is still central to most modern membrane and cell models and there are excellent descriptions in many review articles and even text books . > Figure 3.1 is a schematic circuit diagram that shows the essentials of the Hodgkin–Huxley formalism applied to the cardiac myocyte. Components include an expression for membrane potential expressed as the product of ionic currents and the capacitance of the lipid bilayer that makes up cellular membranes. The original models of Hodgkin and Huxley applied only to nerve axon and included single Na+ and K+ currents; to

represent cardiac action potentials substantial modifications are necessary. The basic equation from the circuit in > Fig. 3.1 that drives all models based on the formalism of Hodgkin and Huxley is

$$\frac{\partial V_{\mathrm{m}}}{\partial t} = -\frac{1}{C_{\mathrm{m}}}(I_{\mathrm{ion}} + I_{\mathrm{stim}}) \qquad (2.1)$$

where Vm is the membrane potential, Cm is the membrane capacitance, Iion is the sum of the active ion currents, and Istim is a stimulus current (usually) required to depolarize the cell to reach threshold. The ionic current is the sum of individual currents that flow through ion-selective channels driven by an equilibrium potential and regulated by a time- and voltage-dependent conductance, expressed as

$$I_{\mathrm{ion}} = I_{\mathrm{Na}} + I_{\mathrm{K}} + I_{\mathrm{Ca}} \ldots \qquad (2.2)$$

where . . . indicates that the list of individual currents is variable and may include different subspecies of currents for the same ion, each having a different time and voltage dependence. Each individual current, in turn, can be expressed according to the circuit diagram as the product of a voltage difference and a conductivity

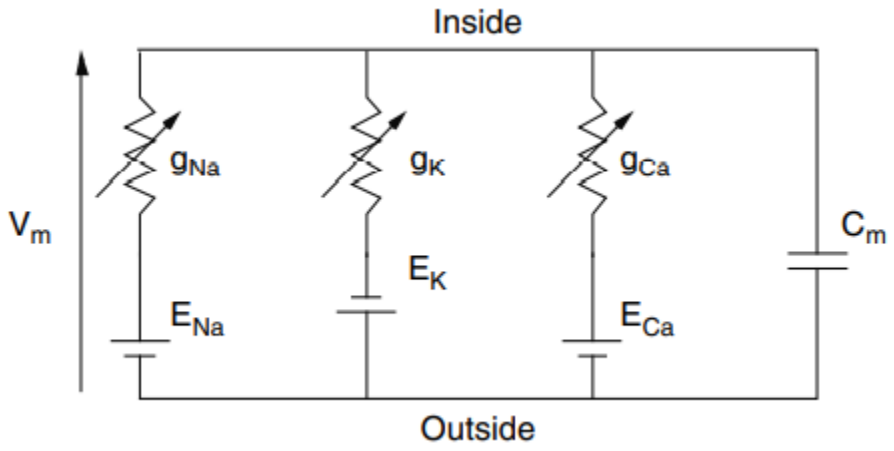

Figure 2-8 Circuit diagram for the Hodgkin–Huxley formalism applied to cardiac myocytes

The parallel circuit of membrane capacitive current and individual ion currents passing through variable resistors captures the essence of the electrical behavior of the cell.

$$i_X(V,t) = (V_m - E_X)g_X(V,t),$$

(2.3)

where IX is the current for ion X, which flows whenever membrane voltage Vm diverges from the equilibrium potential for the particular ion, EX, and the conductivity for that ion, gX, is non-zero. The equilibrium potential, EX, is a function of the relative concentrations of the particular ion on both sides of the semipermeable membrane and is predicted by the Nernst equation:

$$E_X = \frac{RT}{Z_X F}\ln\left(\frac{[X]_e}{[X]_i}\right),$$

(2.4)

where R is the gas constant , T is temperature, ZX is the valance of ion species X, F is the Faraday constant, and [X]e and [X]i are the extracellular and intracellular concentrations respectively, of ion species X. Conceptually, the Nernst potential is the membrane potential difference at which there is no incentive for net movement of a particular ion, i.e., the voltage at which, at least for a particular ion, there is equilibrium between inward and outward currents. At any other membrane potential, each ion will experience a net driving force equal to Vm − EX that will move ions across the membrane if the membrane has adequate permeability to that ion. With this formulation, the descriptions of individual ion currents reside essentially in the time and voltage dependence of the conductivities. As part of their original formulation, Hodgkin and Huxley developed simple expressions for these conductivities based on first-order rate equations that in some way capture their idea of the underlying channel behavior. They fitted the parameters of these equations in order to agree with experimental voltage-clamp results. The resulting formalism remains the most commonly used in modern simulation of electrically active membranes. The first model of cardiac membranes was by Noble et al. in 1962 and there are at least ⋅⋅ different variations on this theme known in the literature today. Each model varies in terms of the number and type of individual ion channels included and the animal species for which individual parameters are known from experimental studies. Perhaps, most commonly used are the models from Luo and Rudy, which include not only ion channels but also support for intracellular buffering of calcium. Sachse has recently summarized very well the state of membrane simulation models, including both electrical and mechanical models. From the perspective of the electrocardiographic forward problem, the choice of membrane simulation model is always sharply constrained by the limits of computation. Because of the desire to simulate pieces of myocardium up to the size of whole heart, it is necessary to discretize the tissue into hundreds of thousands, even millions, of tiny elements, each of which contains its own membrane model. In order to then be able to compute electrical activityfor even a reasonable part of a single heartbeat, it is necessary to find efficient, simplified ion-channel models. This simplification in model comes at a cost of accuracy but is an acceptable

compromise, in part, because the very existence of tissues reduces the impact of individual cells or small groups of cells on the overall behavior of the simulation.

2.3.2. Models of Cardiac Tissue

In this section, we step from the cellular level to the tissue and whole heart and describe models that capture aspects of their electrical behavior that are relevant to the forward problem. As with the cell, some models of myocardium represent a biophysical approach and are based on accurate anatomical features of the heart. Others are parametric in that they use abstractions of some sort to capture the essential features of the electric fields from myocardium, often requiring a certain distance between the source and the observation location. In general, such parametric representations are less accurate but much simpler to compute; their other major weakness is that it is rarely possible to measure them directly or even to uniquely associate their parameters to anatomical or physiological characteristics of the real heart.

2.3.2.1. Discrete Source Models

The simplest and best-known means of representing the electrical activity from the heart is the heart dipole, first described and applied to the heart by Einthoven and still the description with the largest impact on clinical education and the interpretation of the ECG. While simple to grasp physically and to compute numerically, the single dipole is only a barely adequate description of the electrical activity of the heart and even at the body surface cannot explain certain features of potential distributions and the ECG. If anything, it is remarkable that a representation encapsulated by only six parameters can do as well as it does and this fact is perhaps the best evidence of the attenuation, smoothing, and smearing effects of the volume conductor that we explore in more depth in > Sect. 2.5. However, for the purpose of quantitative simulation and the detailed study of almost any aspect of normal or abnormal cardiac electrophysiology, the dipole is an inadequate model and hence is rarely considered in contemporary research studies. There are other forms of discrete sources, based on combinations of dipoles or other higher order forms of this basic current source. Some have proved useful in very special circumstances, such as representing the earliest phase of an ectopic beat, when it resembles a pair of waves moving in opposite directions and hence as a pair of dipoles or a quadrupole. The occurrence of an accessory pathway connecting the atria and ventricles as it arises in the Wolff–Parkinson–White (WPW) syndrome has also been modeled as a discrete dipole that is active only for a few milliseconds. This accessory pathway is a tissue band that links the atria with the ventricles and thus forms a connection secondary to that via the atrio-ventricular (AV) node. This connection can create a circular pathway of excitation that goes from atria to ventricles back to atria and thus a reentrant circuit that can be the substrate for ventricular arrhythmias and even death. During the instant in which the excitation travels through the accessory pathway, this bioelectric source is

very focal and discrete, and the electric field from this source resembles that of a single current dipole. There have also been many reported studies in which a set of distributed dipoles, each with unique orientation and activation times has represented the potentials from the heart – as we will see, this is still a representation that allows models of propagation to generate extracardiac fields. One such model was even the source for a complex inverse solution expressed in terms of on time and off times for a multiple dipole source. In the contemporary context, there are more accurate and sophisticated formulations that are both more accurate, and through modern computers, tractable sources of cardiac bioelectricity. The rest of this section focuses on these, but for more detail on discrete models, the review by Gulrajani is an excellent resource.

2.3.2.2. Bidomain Method

One of the most successful and perhaps initially confusing approaches to representing the electrical activity of the myocardium is the bidomain technique, first conceived by Schmitt, proposed by Miller and Geselowitz and Tung for the heart, and later expanded and used by others to examine all facets of cardiac excitation and stimulation . We present here only a brief overview of the method and refer to an excellent review by Henriquez for more details. The main goal of the bidomain approach is to simplify through a process of "homogenization" the features of an aggregate of individual myocytes so as to enable feasible computations of pieces of heart tissue and ultimately the entire heart. Computation that did not employ such a simplification but instead included every cell of the heart would require approximately 10 billion sets of parameters for each time instant, where each set of parameters could include as many as 30 variables, clearly beyond the scope of any computer. The bidomain approach makes explicit use of the fact that the heart is essentially a syncytium, that is, every cell connects via its immediate neighbors to all other cells. This means that, in principle, stimulating a single cell will eventually lead to all cells firing an action potential (and contracting) in a sequence. The bidomain then describes the heart as composed of two domains, one for all intracellular space and the other for all extracellular space, both coexisting in the same physical space (the myocardium). Thus, what is actually a discrete syncytium of many individual cells, each with their its transmembrane voltage, becomes two continuous domains. Intracellular potential and extracellular potential then become continuous functions of space, as does their difference, the transmembrane voltage. Similarly, other parameters of the tissue, most notably electrical resistance, become continuous functions in the intracellular and extracellular spaces. Joining the two domains of the bidomain is the membrane, which in the classic bidomain has no volume but is likewise distributed everywhere throughout the tissue. Most importantly, the membrane contains the ionic channels represented by voltage- and time-dependent currents that generate action potentials within the cells. > Figure 3.2.1 shows schematically the organization of intracellular and extracellular spaces with a

membrane linking the two. Also visible in the figure is the current that leaves the extracellular space and travels into the extramyocardial volume conductor; it is this current that is ultimately responsible for the torso and body-surface potentials (ECG). As with other aspects of the forward problem, analytical expressions in terms of continuous functions are rarely available so that numerical approaches or discrete approximations of myocardial geometry are necessary. Thus, somewhat paradoxically, what starts as a discrete arrangement of myocytes eventually becomes a discrete organization of small pieces of tissue, each small enough to be treated as uniform with regard to potentials and ionic currents. We describe this process of homogenization and subsequent discretization in the following section. The derivation of the governing equations of bidomain techniques begins with the statement of Ohm' s Law in terms of conductivities for the intracellular and extracellular spaces as

$$\vec{J}_{\mathrm{i}} = -\sigma_{\mathrm{i}} \nabla \phi_i \tag{2.5}$$

$$\vec{J}_{\mathrm{e}} = -\sigma_{\mathrm{e}} \nabla \phi_{\mathrm{e}}, \tag{2.6}$$

where the subscripts i and e on each of $\vec{J}$, σ, and ϕ indicate intracellular or extracellular spaces respectively.

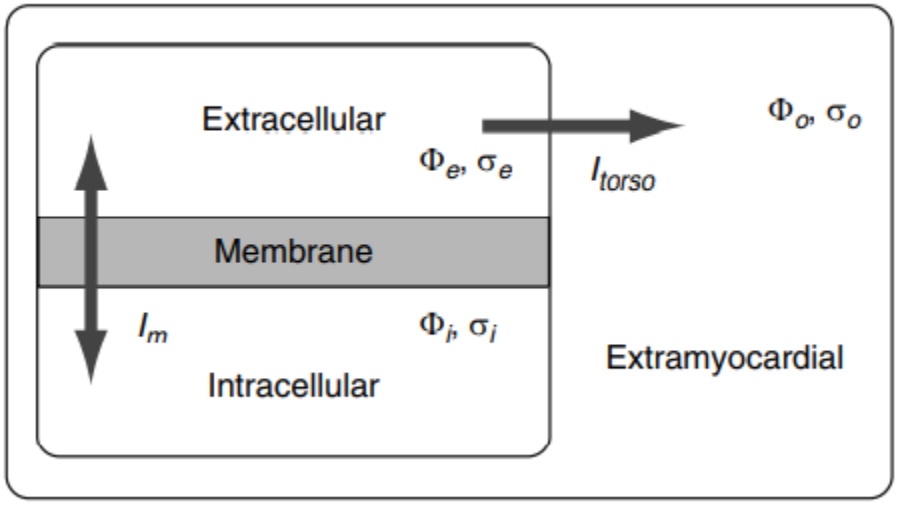

Figure 2-9 Schematic view of the bidomain method of describing myocardium

The heart contains intracellular and extracellular spaces that are linked by a membrane. Currents from the extracellular space (Itorso) flow into the extramyocardial volume conductor to produce body-surface potentials.

Conservation of current requires that whatever current leaving the intracellular domain must enter the extracellular: a condition that is possible because both domains are assumed to exist at the same point in space. This allows one to write

$$\nabla \cdot \vec{J}_{\mathrm{i}} = -\nabla \cdot \vec{J}_{\mathrm{e}} \tag{2.7}$$

Substituting (2.6) into (2.7) provides a conservation equation in terms of the intracellular and extracellular potential fields as

$$\nabla \cdot (\sigma_{\mathrm{i}} \nabla \phi_i) = -\nabla \cdot (\sigma_{\mathrm{e}} \nabla \phi_{\mathrm{e}}) \tag{2.8}$$

We can now make use of the definition of the transmembrane potential

$$V_{\mathrm{m}} = \phi_i - \phi_{\mathrm{e}} \tag{2.9}$$

to express (2.8) in terms of Vm and φe by subtracting ∇·(σi∇φe) from both sides to obtain ∇ · ((σi + σe)

$$\nabla \cdot ((\sigma_{\mathrm{i}} + \sigma_{\mathrm{e}}) \nabla \phi_{\mathrm{e}}) = -\nabla \cdot (\sigma_{\mathrm{i}} \nabla V_{\mathrm{m}}) \tag{2.10}$$

This equation is the first of the two bidomain equations and calculates the extracellular potential field, given a transmembrane potential distribution. Note that this is essentially a form of Poisson' s equation with a source term based on the current density associated with the transmembrane potential field. In fact, the extracellular domain of the bidomain is contiguous with the passive regions outside the heart so that one could consider the whole torso to be a bidomain in which the contributions from Vm are only non-zero in the heart region. As shown in > Fig. 3.2.1, any transfer of current between the intracellular and extracellular domains must pass through the intervening membrane so that

$$\nabla \cdot \vec{J}_{\mathrm{i}} = -\nabla \cdot \vec{J}_{\mathrm{e}} = A_{\mathrm{m}} I_{\mathrm{m}} \tag{2.11}$$

known as the surface-to-volume ratio of the bidomain membrane and essentially describes how much membrane surface area is present per volume of tissue. The function Im describes the current flow across the membrane per unit of membrane area. It is the sum of a time-dependent capacitive current and a second current, Iion, representing the flow of ions through selectively permeable channels in the membrane as described in > Sect. 2.3.1:

$$I_{\mathrm{m}} = C_{\mathrm{m}} \frac{\partial V_{\mathrm{m}}}{\partial t} + I_{\mathrm{ion}}, \tag{2.12}$$

where Cm now denotes the membrane capacitance per unit area. Combining (2.11) and (2.12) gives

$$\nabla \cdot (\sigma_{\mathrm{i}} \nabla \phi_i) = A_{\mathrm{m}} \left(C_{\mathrm{m}} \frac{\partial V_{\mathrm{m}}}{\partial t} + I_{\mathrm{ion}} \right), \tag{2.13}$$

in which Iion is a function of Vm, and therefore, it is common (but not universal) practice to re-state the left hand side of (2.13) in terms of Vm by adding and subtracting ∇·(σi∇φe) and again imposing the definition of Vm to write

$$\nabla \cdot (\sigma_i \nabla V_m) + \nabla \cdot (\sigma_i \nabla \phi_e) = A_m \left(C_m \frac{\partial V_m}{\partial t} + I_{ion} \right) \tag{2.14}$$

Equation (2.14) is the second bidomain equation and from it, one can calculate the transmembrane potential. It is possible for an external stimulus current to be applied to either domain (Is1 and Is2); this allows the two bidomain equations to be written as

$$\nabla \cdot ((\sigma_i + \sigma_e) \nabla \phi_e) = -\nabla \cdot (\sigma_i \nabla V_m) + I_{s1}, \tag{2.15}$$

$$\nabla \cdot (\sigma_i \nabla V_m) + \nabla \cdot (\sigma_i \nabla \phi_e) = A_m \left(C_m \frac{\partial V_m}{\partial t} + I_{ion} \right) - I_{s2}, \tag{2.16}$$

These two equations are highly coupled and indeed solutions can prove to be computationally prohibitive over large domains and/or on high resolution meshes. If either the extracellular domain is assumed to be highly conducting (σe ~ ∞) or the domains are assumed to be equally anisotropic (σi = kσe), it is possible to reduce the bidomain system to a single equation with a significant reduction in computational complexity. The result is known as the monodomain equation,

$$\nabla \cdot (\sigma \nabla V_m) = A_m \left(C_m \frac{\partial V_m}{\partial t} + I_{ion} \right) - I_s, \tag{2.17}$$

In which σ is defined by σ−1 = σ−1 e +σ−1 i .This expression is essentially a nonlinear cable equation when reduced to its onedimensional form. Thus one can think of the bidomain approach as a multidimensional generalization of the nonlinear cable equation

2.4. Simulating Propagation

To simulate a full heartbeat, it is necessary to have bioelectric sources that vary with time, ideally over more than one beat if the goal is to study arrhythmias or other effects related to heart rate. The most realistic way to generate time-varying sources is to simulate the spread of excitation within tissue, and there are many approaches to describing such propagation in the heart. As with all simulation components, there exists a range of levels of sophistication and degrees of realism and flexibility that such a model can simulate. There are two stages to this type of simulation and multiple approaches to both. The first step is

to predict the sequence in which different regions of the tissue depolarize, the actual spread of excitation. The result is a set of activation and/or repolarization times, parameters which are immediately useful to address some highly relevant questions. The second step is to use the timing values of the excitation and repolarization to generation potentials either within the tissue, or more often, on the epicardium and endocardium. In the context of the forward problem, one usually wishes to generate such potentials at the body surface (although we leave this last step to the section on volume conductors). As described in > Sect. 2.3.2, a common approach to calculating potentials is to represent different regions of the heart as dipole sources and use the activation wavefronts to provide their orientation and active/repolarization times and thus determine their timing. Another approach, described in detail in > Chaps. 6 and > 7, is to represent the sources as moving double layers, with timing and path determined from the propagation parameters. Rather than explicitly computing the spread of excitation, it is also possible to assume knowledge of this sequence, e.g., to acquire it from experiments, and then generate extracardiac potentials from this information. We use this approximate taxonomy of stages to describe and categorize specific models and begin with simulation approaches that generate the sequence of activation or spread of excitation, which we take as equivalent expressions. Here, again, we emphasize the intuition of each approach over its mathematical description and solution and refer readers to other sources for more detail.

2.4.1. Physiology Background

We begin with a qualitative description of the process by which excitation travel through the heart. Like all organs, the heart comprises a large number of cells, each of which must be stimulated (partially depolarized) in order to generate an action potential. At the tissue level, the question arises as to how the impetus to depolarize passes from one cell to the next. The answer is via Ohm' s Law and gap junctions. Ohm' s Law (I = V/R) states that current will flow when a potential difference arises and the resistance between regions at different potentials is low enough. Thus, if one cell depolarizes, it will become more positive than its neighbors so that current will flow, providing there is a conductive pathway. The conductive pathway is provided by what are known as "gap junctions". Gap junctions are a form of protein channel embedded in the cell membrane that form direct connections between the interior of one cell and the interior of a neighboring cell through which charged ions may flow – each one represents a resistor with relatively low resistance compared with surrounding tissue. While return currents do flow in the extracellular space, the gap junction connections are the primary means of transferring electrical information between cardiac cells. > Figure 2.2 in > Sect. 2.2.2 shows schematically the

relationship between two cells coupled by gap junctions, the primary current pathway through the gap junctions, and the return current through the extracellular space. There are some characteristic features of

propagation in myocardium that follow from these mechanisms and any attempt at simulating propagation must also take them into account. In fact, appreciation of these features can lead to simplified formulations of propagation, as we will show. The first notable feature is syncytial nature of the heart described in > Sect. 2.3.2, by which each cell is connected to all other cells by means of a series of neighbors – a single depolarized cell will eventually cause all cells to depolarize as long as there is a path to each cell. Similarly, all paths from one cell to a non-neighboring cell must go through intermediary cells, i.e., there are generally no short cuts or direct connections between remote regions of the heart. The exception to this rule is a set of special cells, superimposed and only partially connected to the rest of the heart, that are typically noncontractile and preferentially carry extrication more rapidly than surrounding tissues. This conduction system is responsible for the coordination and timing of the spread of excitation, ensuring, for example, through a branching network of fibers that most of the subendocardial regions of the ventricular septum are stimulated almost simultaneously. Not all elements of the conduction system accelerate the spread of excitation. In fact, the atrio-ventricular (AV) node is the only electrical link between atria and ventricles – in a normal heart, at least – and it exhibits especially slow conduction speed because of weak intercellular coupling and smaller, rather slowly rising action potentials of the cells. A complete propagation model of the heart must incorporate all these behaviors, or at least must be able to mimic them in some form. We will now describe some of the more common forms of simulating propagation in physical and mathematical terms and also direct the reader to > Chap. 6 of this volume for a discussion of the specific example of double layer sources.

2.4.2. Cellular Automata

The concept of cellular automata dates to the 1940s and two scientists, Stanislas Ulam and John von Neumann, but by far the most widely recognized use is John Conway' s Game of Life. Moe et al. adopted this approach in the early 1960s and the methods still find great popularity today due to its computational efficiency. A cellular automata approach divides the electrical wave propagation problem into two components. First, the domain of interest is divided up into a regular grid. Each point on the grid has a set of neighboring points, the number of which depends on the dimension of the problem and the topology of the grid. The second component is the automaton that is used to represent the behavior of a single cell and is so named because its actions are solely determined by a set of internal rules. Note that because of computational constraints, an element or cell of the grid in this context is usually substantially larger than a physical cardiac cell. In its simplest form, a cell in a cellular automata model would have two states, resting and excited (essentially on and off) plus a set of rules to describe the transition from one state to the other and back again. This system mimics a piecewise approximation of an action potential. To then replicate the behavior necessary for propagation, there are rules that determine how the state of one cell affects

those of its neighbors. Typically, a cell may transition to the excited state a fixed time after it senses that one or more of its neighbors have been excited. The same cell would then have a rule whereby it returns to the resting state after some further period of time has elapsed, thus encapsulating the ordered cell-by-cell coordination prescribed by the underlying physiology. More complex – and realistic – cellular automata incorporate additional states in order to capture, for example, refractory properties. One example of such a model from Bailie et al. includes three states: quiescent (Q), excited (E), or refractory (R). In addition, there are a total of six rules that govern the transition into and out of each state, summarized in > Table 2.1

starting from a quiescent state, a cell makes the transition to the excited state if it detects that one or more of its neighbors are in an excited state. Next, after a prescribed time period, the cell transitions to the refractory state, and then after another period of time, returns to the initial quiescent state. > Figure 2.10 shows the initial stages of propagation using this simple automaton.

State(t)	Count(t)	Neighbor(t)	State(t+1)	Count(t+1)
Q	(any)	0	Q	0
Q	(any)	≥ 1	E	1
E	$< E_T$	(any)	E	Count(t)+1
E	E_T	(any)	R	1
R	$< R_T$	(any)	R	Count(t)+1
R	R_T	(any)	Q	1

Table 2.1

Figure 2-10 Propagation from a point source using a cellular automata approach

Initially, one cell is set to be in an excited state (light) while the remainder are quiescent (dark). Subsequent images show neighboring cells making the transition to the excited state

Leon et al. described in a series of studies a cellular automata model that incorporated anisotropic propagation in a human ventricle model by altering the rules of interaction among cells. In order to change from resting to activated state, each target cell received inputs from all neighboring cells, each with its contribution weighted by the direction between the target cell and each neighbor relative to the local fiber direction. In this way, neighboring cells that lie along the local fiber direction have more influence on a

target cell than neighbors that lie in the cross-fiber direction. They then used this model to simulate reentrant propagation and to illustrate some of the effects of rotational anisotropy on conduction in a very realistic heart model with 1 mm resolution. Wei et al. constructed an anatomically based heart mesh that included descriptions of the atria and ventricles, using approximately 50000 connected elements. The structure was such that each element was connected to 12 adjacent neighbors. The authors divided the heart into 16 cell types with different properties and simulated a normal cardiac excitation pattern. At each time step 54 dipolar sources were constructed from the transmembrane potential distribution, and these were used to simulate the ECG and also the vectorcardiogram. WeiXue et al. used CT images to construct a 65,000-element heart model at a resolution of 1.5 mm. The hexahedral elements were arranged in a cubic closepacked structure meaning that each was connected to 26 neighbors. Again, a cellular automata approach was used to model propagation, and, in this case, the primary interest was in the epicardial, and body surface potential distributions associated with Wolff–Parkinson–White syndrome. Hren et al. developed even more finely resolved cellular automata models with 0.5 mm cell size and used it to simulate various forms of arrhythmia and determine a functional limit of the resolving power of body surface potentials, and to investigate their use in guiding radio-frequency ablation procedures. Following the success of the Visible Human Project, there has been much interest in using this high-resolution data source in biomedical computations. Sachse et al. segmented the cardiac geometry from the Visible Man, and like Weixue et al., used the pixelated structure as a template for a cellular automata solution. In their study, the transmembrane potential distribution at each time point was used to construct ECG signals from both normal sinoatrial rhythm and also a number of conduction pathologies. Freudenberg et al. also constructed a heart model but, in this case, using the Visible Female data set. The resolution of the original images was scaled down by a factor of two in each direction to limit the number of variables that the computation had to encompass. The main advantage of the cellular automata approach over more biophysically detailed models is the computational speed; whole heart simulations are possible without the need for expensive supercomputing facilities. There are, however, drawbacks to this approach. For most implementations, the shape of the resulting wavefront is dependent on the topology of the grid – a mesh tiled with squares produces a topologically square wavefront). Although it is possible to generate curved wavefronts, these are restricted by the need for a characteristic dimension for the curvature. The cellular automata approach is also unsuitable for simulations in which dynamic cellular activity is important, such as the onset of ischemia, as they do not include realistic cellular electrophysiology and hence cannot respond to most external stimuli

2.4.3. Computing Extracardiac Potentials

With a propagation sequence available, either from measurements or from one of the propagation models already described, the second stage of bioelectric source modeling is converting information that is essentially timing into the electric sources – typically currents – that would then generate the potentials measured from the heart or the body surface. The typical approach is to use the activation time and some assumptions about the resulting potential gradients that arise at the spreading wavefront to create some reasonable number of current sources. Because current and current density have direction, it is also necessary to determine the orientation of discrete current sources, which in turn requires knowledge of the spatial orientation of the spreading activation wavefront.The final component required is the amplitude of the current sources, which one can derive naturally from the spatial gradient of the potential in the myocardium. The typical assumption during the activation phase is that the most significant potential gradients are localized in the region of the spreading wave of excitation so that it is necessary to evaluate only those regions of the heart. One particular simplifying assumption is to assign a fixed potential difference across all regions of the activation wavefront, which assumes that both the extracellular and intracellular potentials are identical for every cell. A more realistic version of this scheme assigns an action potential to each region of the heart, based either on a local implementation of one of the cell simulation models described in > Sect. 2.3.1 or on measurements and taking into account the natural variations in action potential characteristics over the heart. Mathematically, one can express this approach as computing a current density from Ohm' s Law

$$\mathbf{J}_i = -\sigma_i \nabla \Phi_i,$$

(2.27)

where Ji is the current density in some region i, σi is the effective intracellular conductivity (which may be a tensor quantity), and ∇Φi is the potential gradient over the same region. Depending on the model geometry and structure, there is a discrete form of this equation that applies at some appropriate resolution either throughout the heart or in regions of appreciable gradients (defined, for example, by the location of the activation wavefront at any given time). In some formulations of the extracardiac potential, this expression for current density is enough, while in others, a more discrete form is necessary and one computes a dipole moment by integrating the current density over some region in space,

$$\mathbf{P}_{l,m,n} = -\int \sigma_i \nabla \Phi_i \, dl \, dm \, dn,$$

(2.28)

Where dl dm anddn describe some region in the representation of the region collapsed into a single dipole with moment $P_{l,m,n}$

2.4.4. Applications

The applications of the bidomain approach and propagation modeling are very numerous and include a wide range of problems. One of the earliest applications of this approach to cardiac propagation was by Henriquez et al., who constructed detailed models of bundles of cardiac tissue and examined the critical values for spread of excitation [].A more modern version of this approach has recently appeared in simulations by Stinstra et al., first of the passive electrical characteristics of cardiac tissue [], and then using those parameters to simulate at unparalleled spatial resolution the effects of cell shape and size on propagation. Henriquez and his various colleagues have published a series of report of applying bidomain approaches to ever more complex slabs of tissue to investigate the role of structure, coupling and especially anisotropy on cardiac excitation in the ventricles. Bidomain models have also become tractable enough to simulate some behaviors in whole heart models. For example, Fischer et al. used a bidomain model of the heart together with a propagation model to simulate cardiac potentials under realistic conditions of anisotropy []. Other studies by Weinstein and Henriquez et al. have described the first computations of cardiac potentials in mice using bidomain models .Hopenfeld et al. also used a realistic, anisotropic dog heart model and the bidomain approach to simulate cardiac potentials from acute ischemia under varying conditions of ischemic zone location and depth []. The fact that they simulated an essentially static condition from potentials during the plateau phase of the action potential made these simulations very tractable compared with models with propagation. For that, the model allowed unprecedented examination of the relationship between ischemic border zone shape and alignment with local myocardial fiber structure. There are some less obvious applications of the bidomain for problems of tissue stimulation that make use of the unique ability of this formulation to capture spatially variable anisotropy. In one such application, Frazier et al. used a bidomain model of heart tissue to investigate the fields that arise during extracellular stimulation of tissue and thus to investigate the role of tissue structure, electrode location, and stimulus strength on stimulation . Using even stronger extracellular fields, Trayanova and her colleagues have published many studies on defibrillation using the bidomain approach . Recent studies by Jolley and Triedman et al. have illustrated the feasibility of carrying out simulations of cardiac defibrillation using realistic electrode placements inside the human child thorax. Successful studies such as these will make it possible to simulate intracardiac defibrillator placement in subject-specific models and thus optimize lead placement and defibrillation protocols for each patient.

2.5. Models of Volume Conductor Potentials

In this section we describe the mathematical models and numerical methods required to compute body surface potentials from cardiac bioelectric sources. This is the last of the four components of the complete forward solution introduced in > Sect. 2.3

, which, in essence, describes the biophysical characteristics of the volume conductor, a passive medium with heterogeneous electrical conductivity. In practice, however, it is necessary to match the model of the volume conductor to the specific representation of the source and the desired outcome of the forward problem. For example, if one represents the heart as Einthoven did as a single current dipole and seeks to compute the body-surface ECG, the volume conductor model and the associated mathematical formulation will differ from the case of the source described in terms of activation times on the epicardial and endocardial surfaces. Thus, any discussion of the volume conductor potentials is closely tied to the source formulation, a fact that we reflect in the organization of this section. The choice of numerical method is also a product of the type of source model and desired outcome of the simulation, as well as assumptions about the volume conductor itself. For example, a forward problem based on epicardial and endocardial surface activation times and under assumptions of heterogeneous but isotropic torso volume conductor leads naturally to a numerical implementation based on the boundary element method (BEM). By contrast, a solution based, for example, on bidomain computations of cardiac currents embedded in an anisotropic model of the heart and heterogeneous (and possibly anisotropic) torso is better served by an implementation using the finite element method (FEM). Wherever possible, we will indicate the numerical approach best suited for a particular problem formulation and direct the reader to > Sect. 2.6 for details of the associated numerical methods. This section describes the biophysics and mathematics of volume conductors required to form the bridge between the descriptions of bioelectric sources in the previous sections to the numerical methods that are necessary to implement working solutions to electrocardiographic forward problems. It is impossible in a single chapter to provide comprehensive coverage of all combinations of source models and volume conductors so that we focus here on some of the more frequently studied scenarios and describe at least the main numerical methods in contemporary use.

the physics of potential fields describes this component of the forward problem of this volume

and arrive at a more general framework, specifically one that supports a wider range of realistic tissue characteristics. The basic formulation of all such problems leads either to Poisson' s or Laplace' s equation, i.e.

$$\nabla \cdot \left(\sigma \nabla \Phi\right) = -i_v = \nabla \cdot \vec{J}^i, \tag{2.29}$$

the general expression that applies to inhomogeneous, anisotropic conductivity. In case the conductivity is anisotropic, σ denotes the tensor character

Similarly, we can write a Laplace' s equation

$$\nabla\cdot(\sigma\nabla\Phi) = 0 \tag{2.30}$$

for any region in which there are no active sources, i.e.,

$$iv = 0.$$

To make use of Laplace' s equation, the solution domain cannot include the sources, which then appear as imposed boundary conditions of the resulting solution. These boundary conditions include electric potentials that are known at some subset of the problem domain, typically on or within the heart, and this is known as the "Dirichlet" boundary condition. The solutions of (2.29) and (2.30) also depend on boundary conditions that arise from the physics of the problem. At the boundary of the outer surface of the torso and air, for example, no current flow leaves the body so that

$$(\sigma\nabla\Phi) \cdot n = 0, \tag{2.31}$$

where n is a unit vector pointing in normal direction to the surface. This is known as the "Neumann" boundary condition on the flux through the surface. At the boundaries within the volume conductor there are also boundary conditions dictated by the fact that potentials must be equal on both sides of a shared boundary and normal current flow must therefore also be the same, i.e., for a boundary between compartments p and q

$$\Phi p = \Phi q, \tag{2.32}$$

$$(\sigma p\nabla\Phi p) \cdot np = (\sigma q\nabla\Phi q) \cdot nq \tag{2.33}$$

2.5.1. Analytical Models

Under highly simplified conditions, it is possible to formulate closed form or analytical solutions to (2.29) and (2.30) and there is a rich history of applying this approach to the forward problem in electrocardiography. The simplified conditions typically include a source described in terms of a modest number of discrete current dipoles and a volume conductor with a mathematically simple shape, for example a circle, sphere, or cylinder. Examples of this approach include studies from Wilson and Bayley [] and then Frank [], who computed body-surface potentials from a single current dipole source located in a

spherical surface model of the torso. Burger et al. and Okada carried out similar studies using cylindrical model of the torso and concentric and eccentric dipole sources, which led eventually to the extensive studies based on a model of eccentric spheres developed by Bayley and Berry. Rudy et al. later further developed this approach extensively ,and it is still in use, largely as a means of validating solution techniques based on numerical approximations of the sources and volume conductors (as we describe in the following sections). It is rare to use analytical models for any realistic solutions to forward (or inverse) problems in electrocardiography, simply because the resulting accuracy is inadequate when compared with measured values in a real subject/patient.

2.5.2. Discrete Source Models

There is a rich history of forward problem formulations based on discrete source models of the heart, The underlying notion of all these approaches is that a very simple current source can adequately represent the electrical activity of the heart, at least when viewed from some distance from the heart, e.g., the body surface. The huge advantage of these approaches is the small number of parameters required to describe the sources and the resulting simplicity of the associated forward (and inverse) problems .The most notable disadvantage is that there are no clear physical or physiological links between discrete sources and either the true sources they represent or any directly measurable quantities; there is no way to directly measure dipole locations or amplitudes nor to unambiguously link their parameters to action potentials or spread of activation. Discrete source models nevertheless play important roles in both the analytical approaches already outlined and the numerical approaches described in the following section. The descriptions of bioelectric fields from discrete sources begin with the expression for the potential from a single dipole in an infinite space, rewritten as

$$\Phi(\vec{r}') = \frac{1}{4\pi\sigma}\frac{\dot{R}}{R^3}\cdot\vec{P}(\vec{r}), \tag{2.34}$$

where $\vec{R}$ is the vector from the location of the dipole to the location $\vec{r}'$ in space where Φ is calculated and $\vec{P}(\vec{r})$ is the strength (dipole moment) of the dipole at location $\vec{r}$, specifying both the direction and the magnitude of the current dipole source. To compute the potentials from a current dipole located within a finite medium, even one with several regions having different conductivities (σ) is also possible using the equation

$$\Phi(\vec{r}') = \frac{1}{4\pi\sigma}\frac{\vec{R}}{R^3}\cdot\vec{P}(\vec{r}) - \frac{1}{4\pi}\sum_{l=0}^{N_S}\int_{S_l}(\sigma_l^+ - \sigma_l^-)\Phi(\vec{r})\,d\Omega_{rr\prime}, \tag{2.35}$$

where σ+ l and σ− l are the conductivities on either side of the boundary l and dΩrr′ is the solid angle . With a numerical implementation of this equation (usually via the boundary element method described in the following section) it is then possible to compute the potential at any point provided the dipole location, orientation, and strength and the spatial locations of all boundaries and the conductivities around them are known. It is also possible to extend this approach to other discrete sources, i.e., quadropoles or octopoles, which are capable of representing more complex versions of the cardiac sources at the cost of more parameters but still without a clear link to measurements or physiological sources [1]

2.5.3. Cardiac Surface Potential Models

This form of the forward problem is one of the two dominant approaches in modern applications and is based on a source description of epicardial potentials. More precisely, the potentials on any closed surface that encompasses all active cardiac sources forms a completely equivalent representation of the true cardiac sources. Out of convenience – because it permits direct measurement of the source parameters – one usually selects the enclosing surface to be the epicardium (or pericardium) and the challenge then becomes to derive a mathematical formulation that predicts the potentials throughout the thorax (include the body surface) from this source. Barr and his colleagues solved this problem in the mid-1970s in a way that also provides a natural numerical solution by means of the boundary element method .

Figure> 2.11 describes schematically the configuration and the associated variables involved in linking the potentials the heart surface SH and the body surface SB. In this figure, the region between the heart surface and the body surface is taken to have a uniform conductivity. The derivation of this link as derived from the BEM is described in implies that inhomogeneities in the region between SH and SB, like those of the lungs, can also be treated by means of the BEM. In its numerical form, the transfer can be described by a matrix A. Any element a_{BH} of this matrix represents the effect of the potential at location PH on the heart surface SH to points PB on the body surface SB. The transfer coefficients relate to the specification of the geometry and conductivities of the problem, i.e., it depends only on the volume conductor and not on the time-varying source potentials. Although the formalities of all the terms of the transfer are crucial when carrying out the calculations, it is their qualitative meanings that are fundamental to understanding this entire approach. The equations (and diagrams)

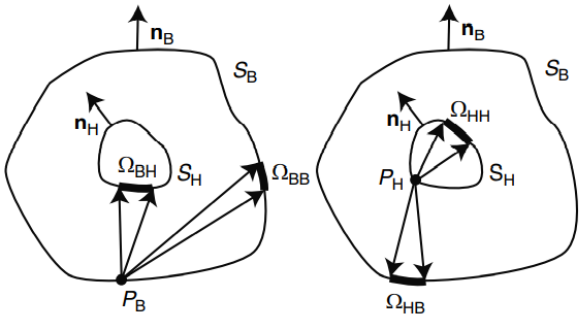

Figure 2-11 Schematic diagram of the forward problem in terms of surface potentials

Schematic diagram of the forward problem in terms of surface potentials. (a and b) represent the approach to calculating the potentials on each of the heart (PB on SB) and body (PH on SH) surfaces respectively. Both surfaces have outward normals (nH and nB) and all active sources are contained within SH. The heavy line segments between the tips of the arrows represent the contribution of a surface element to the potential at the common origin of the arrows and the Ω symbols represent the solid angle that weights each of these contributions.

Reveal that the potential at any point on the surface of, for example, the body surface is the sum of contributions from the potentials (and perhaps potential gradients) at the other surfaces in the problem. The full expression of the forward transfer reads

$$\mathbf{\Phi}_B = \boldsymbol{A}_{BH}\mathbf{\Phi}_H \tag{2.36}$$

The computation of potentials on the heart surface from body surface potentials is based on evaluationg the inverse of the transfer matrix involved, a topic treated in > Sect. 9.3 of > Chap. 9. One of the essential aspects of this approach to the forward problem is that it is based on epicardial (or pericardial) surface potentials, which are, in principle, measurable quantities. In fact, as we outlined in > Sect. 2.2.2, these quantities are measured in experiments and in clinical practice, because they convey directly information about the electrical state of the heart. This form of the forward solution is also unique in the sense that for any given set of source potentials there is one and only one set of associated body surface potentials (as we shall see in the next chapter, the reverse is also true, at least under assumptions of perfect numerical calculations and noise-free measurements). The information necessary to solve this problem consists of a description of the heart and body surfaces and of any other surfaces that separate regions of differing conductivity along with the tissue conductivities of each of these regions.

2.5.4. Equivalent Double Layer Based Models

A final formulation of the forward problem is based on a representation of the spread of activation and repolarization in the heart, and this represents the sources not as variable potential amplitudes but on the timing of activation and repolarization on the epicardial and endocardial surfaces of the heart contain descriptions of these sources (the uniform and equivalent double layers) and the resulting forward problem respectively. The ECGSIM software described in > Sect. 5.6.6, is a very user-friendly and powerful implementation of this forward model.

2.6. Numerical Approaches and Computational Aspects

2.6.1. Introduction

This section describes a family of numerical methods that are common to many different types of electrocardiographic forward problems and in some cases, several of them may solve the same problem. Hence they are combined here into one section rather than organized within the previous sections describing the types of problems to which one might apply them. These methods are common not only to forward problems but also to many other applications across the broadest spectrum of engineering and science, which has the considerable advantage that they are well supported with robust theory and easily available implementations as computer programs and libraries. Here we provide only outlines of these methods with emphasis on the intuition that motivates each and the specific applications to the forward problem.

Please note that in this section the symbol Ω denotes a region in space, rather than a solid angle as used at several other places in this volume, and that Γ denotes its boundary. To provide a common framework for all these methods, we begin with Laplace' s equation (2.30) and note that it is a differential form, also known as a strong form

$$\nabla \cdot (\sigma \nabla \Phi) = 0 \tag{2.37}$$

Numerical solutions to differential equations always involve some degree of approximation and uncertainty. Thus, any numerical solution for (2.37) will also be to some extent approximate. If we use the notation $\varphi\hat{}$ to represent the numerical solution, then the substitution of $\varphi\hat{}$ into the left hand side of (2.37) will not yield identically 0 ; there will be some error, or residual, remaining that one would like to minimize, thereby yielding an accurate answer. Different numerical techniques seek to minimize this residual in different ways, typically by trying to force the average (or weighted) error to be zero. The general expression for this is the weighted residual expression or "weak form" given by

$$\int_{\Omega} \nabla \cdot (\sigma \nabla \phi)\, \omega \, d\Omega = 0, \tag{2.38}$$

In terms of the numerical techniques outlined in the subsequent sections, the finite element method (FEM) and the boundary element method (BEM) both solve the weak form while the finite difference method (FDM) solves the strong form. Consequently the FEM and BEM enforce the governing equation in an integral sense, i.e., the governing equation will be satisfied over each element but not necessarily at any given point. Finite differences, on the other hand, do satisfy the governing equation at the difference points but not necessarily within the space between these points. The choice of numerical approach is closely tied to the specific forward problem and there is no single best method for all applications . The finite difference method (FDM) is intuitively the most straightforward, because it directly estimates the derivatives explicit in the strong form of the equation (Poisson' s or Laplace' s). However, the FDM usually requires a regularly sampled, orthogonal grid for the solution domain, which is rarely the most efficient way to describe complex geometries as they arise in the body. The FDM method requires this grid to encompass the entire volume and can represent anisotropic characteristics of this domain. As a result, the FDM approach has seen widespread use in models of regular slabs of myocardium and also to represent the whole torso .Recent studies have even suggested a generalized FDM approach suitable for irregular grids applied to the bidomain method. The boundary element method (BEM) is based on a grid of the boundaries of a region, e.g., the surfaces that surround the heart, torso, and internal organs of differing electrical conductivity. The BEM is well suited to inhomogeneous tissues but not to regions with anisotropy. Its major advantage lies in the requirement only for surface descriptions, which is often the first step in creating volume grids and is more flexible and easily adjustable than finite element (or finite volume) methods. A related advantage comes from the fact that the number of nodes – the degrees of freedom – involved in the computation is smaller for BEM than FEM (surface rather than volume). As a result, the matrices involved in solving the forward problem by the BEM are generally much smaller than in the other methods, although, the latter involve sparse matrices so that the overall size of the computation may ultimately be comparable. Applications of the boundary element method to forward problems in electrocardiography abound and virtually all the early simulation studies (as well as many contemporary studies) have made use of this technique The finite element method (FEM) and the related finite volume method (FVM) have gained broad support in the past decade, in large part for their general utility, their ability to include all forms of tissue conductivity, and the advanced support available in the form of efficient computational implementations. The chief disadvantage of the FEM is the requirement for carefully constructed volumetric meshes of nodes and polygons: a task that is still arguably the largest barrier to practical use of all forms of numerical methods for partial differential equations. Any changes in

geometry of the heart or torso also require a reorganization of the mesh, often best achieved by starting again from the beginning (or at least from the adjusted surfaces)

2.6.2. The Finite Difference Method (FDM)

The finite difference method represents arguably the simplest approach to solving the generalized Laplace equation governing the potential fields and current flows within the torso volume conductor. The inherent simplicity of the FDM relies on an underlying mesh that is a grid of evenly spaced solution points distributed along orthogonal axes across the solution domain. Meshes for the FDM are relatively easy to construct, at least for pieces of tissue that represent rectilinear slabs or in cases for which image data with equal sampling in all directions are available. The resulting system of equations has some advantageous structure that can facilitate computational solution approaches. The FDM solves the strong or differential form of the generalized Laplace equation shown in (2.37)

2.6.3. The Finite Element Method (FDM)

The finite element method has become increasingly pervasive in traditional engineering disciplines and is also becoming a popular solution methodology for problems in cardiac bioelectricity. The starting equation for the finite element method is the weighted residual form of the generalized Laplace equation given in (2.38), to which we apply Green' s first formula, which can be thought of as a multidimensional version of integration by parts

Given scalar fields f and g with a tensor k, Green' s first formula can be written as

$$\int_{\Omega} \left(f\nabla \cdot (k\nabla g) + \nabla f \cdot (k\nabla g) \right) d\Omega = \int_{\Gamma} fk\nabla g \cdot \boldsymbol{n}\, d\Gamma$$

(2.48)

This also known as the Green–Gauss theorem because the scalar fields must obey the divergence theorem of Gauss. Applying (2.48) to (2.38) yields

$$\int_{\Omega} \nabla \cdot (\sigma\nabla\phi)\, \omega d\Omega = \int_{\Omega} (\sigma\nabla\phi) \cdot \nabla\omega d\Omega - \int_{\Gamma} (\sigma\nabla\phi) \cdot \boldsymbol{n}\omega d\Gamma = 0,$$

(2.49)

And from 2.38:

$$\int_{\Omega} (\sigma \nabla \phi) \cdot \nabla \omega d\Omega = \int_{\Gamma} (\sigma \nabla \phi) \cdot n \omega d\Gamma \tag{2.50}$$

To solve (2.50) using finite elements, as the name suggests, the solution domain is first subdivided into L smaller domains or elements. These elements are nonoverlapping and in combination completely cover Ω. This can be written mathematically as

$$\Omega = \bigcup_{l=1}^{L} \Omega_l \tag{2.51}$$

Typically the geometry of these elements is defined by a set of nodes which, in the simplest cases, correspond to the vertices of the elements. The set of nodes associated with a particular element are known as local element nodes. Field variables can be interpolated between the local element nodes and hence over an element by what are known as basis functions. Instead of performing the interpolation in global space, it is preferable to introduce a local element-based coordinate system referred to here as local element or ξ space. In this space, the axes are orthogonal and the ξ directions are normalized to lie between · and · in each direction. This approach has the advantage that regardless of their physical geometry, each element in the mesh appears identical in ξ space and thus evaluating (·. · ·) over one element easily extends to the remaining elements. The relationship between global and local element space is illustrated in > Fig. 2.12.

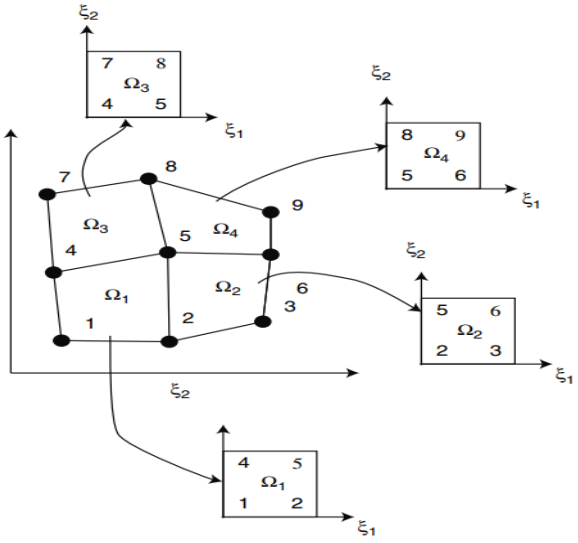

Figure 2-12 A two-dimensional example of the mapping between the solution domain and local finite elelment space

A two-dimensional example of the mapping between the solution domain Ω and local finite element space. The solution domain is divided into four finite elements, Ω ·, Ω ·, Ω · and Ω · with nine global nodes. Each of these elements is then mapped to the normalized local ξ space with four local nodes and the geometric basis functions are used to interpolate nodal quantities over these elements

The details of the basis functions required to interpolate among the local element nodes in ξ space largely depend on the shape of each element and the number of nodes it contains (e.g., a triangular element will have three nodes while a hexahedral element will have eight). As an example, the bilinear interpolation of a variable u between the four nodes of a quadrilateral element can be described by

$$u(\xi_1, \xi_2) = \Psi_1(\xi_1, \xi_2)\, u_1 + \Psi_2(\xi_1, \xi_2)\, u_2 + \Psi_3(\xi_1, \xi_2)\, u_3 + \Psi_4(\xi_1, \xi_2)\, u_4, \tag{2.52}$$

where ξ1 and ξ2 represent the coordinate axes in local element space and both lie between 0 and 1 inclusive. The terms u ·, u ·, u · and u · refer to the value of u at each of the four local nodes, and the four basis functions from (2.52) can be written in terms of ξ1 and ξ2 as

$$\begin{aligned} \Psi_1(\xi_1, \xi_2) &= (1 - \xi_1)(1 - \xi_2) \\ \Psi_2(\xi_1, \xi_2) &= \xi_1(1 - \xi_2) \\ \Psi_3(\xi_1, \xi_2) &= (1 - \xi_1)\xi_2 \\ \Psi_4(\xi_1, \xi_2) &= \xi_1\xi_2 \end{aligned} \tag{2.53}$$

$$u(\boldsymbol{\xi}) = \Psi_n(\boldsymbol{\xi})\, u_n, \tag{2.54}$$

where n represents the appropriate number of local element nodes and noting the vector form of ξ indicating the presence of an appropriate number of ξ coordinates. Using these ideas we can rewrite the volume integral from (2.50) in terms of individual elements and their local ξ coordinates as

$$\sum_{l=1}^{L} \int_{\Omega_l} (\sigma \nabla \phi) \cdot \nabla \omega \, d\Omega_l = \sum_{l=1}^{L} \int_{\mathbf{0}}^{\mathbf{1}} (\sigma \nabla \phi) \cdot \nabla \omega \, |J(\boldsymbol{\xi})| \, d\boldsymbol{\xi}, \tag{2.55}$$

where J is the Jacobian for the transformation from global to local element coordinates and all of the terms inside the resulting integral are expressed in terms of local element space. The emboldened integral notation is used to represent the presence of the appropriate number of integrals – three for a volume

integral over a three-dimensional element. To simplify the expressions in the following step, we make use of the Einstein summation convention, which states that the letter used for an index is arbitrary, a repeated index indicates a summation, and no index may be repeated more than twice. An example of such a summation is as follows.

$$x_i y_i = \sum_{i=1}^{3} x_i y_i = x_1 y_1 + x_2 y_2 + x_3 y_3 \tag{2.56}$$

Within the integral the two gradient terms may be expressed as

$$\nabla \phi = \frac{\partial \phi}{\partial \xi_k} \frac{\partial \xi_k}{\partial x_j}, \tag{2.57}$$

$$\nabla \omega = \frac{\partial \omega}{\partial \xi_h} \frac{\partial \xi_h}{\partial x_j}, \tag{2.58}$$

which is essentially just applying the chain rule for partial derivatives. For a scalar conductivity value the integrand then becomes

$$(\sigma \nabla \phi) \cdot \nabla \omega = \sigma \frac{\partial \phi}{\partial \xi_k} \frac{\partial \xi_k}{\partial x_j} \frac{\partial \omega}{\partial \xi_h} \frac{\partial \xi_h}{\partial x_j}, \tag{2.59}$$

noting the summations over h, j, and k. For information on the extension of this formula to a general conductivity tensor, please see Pullan et al.

Within each element, the dependent variable to be evaluated (in this case φ) can be described via interpolation using basis functions and (2.54), i.e., φ = Ψn (ξ) φn. It is now necessary to also choose a set of weighting functions ω and a common choice are the dependent variable basis functions, i.e., ω = Ψm (ξ), yielding what is known as a Galerkin finite element formulation. Equation (2.59) then becomes

$$(\sigma \nabla \phi) \cdot \nabla \omega = \phi_n \, \sigma \frac{\partial \Psi_n(\xi)}{\partial \xi_k} \frac{\partial \xi_k}{\partial x_j} \frac{\partial \Psi_m(\xi)}{\partial \xi_h} \frac{\partial \xi_h}{\partial x_j} \tag{2.60}$$

he φn terms are now no longer functions of space but instead refer specifically to the value of φ at the n local element nodes. Thus, they can be moved outside the integral

$$\sum_{l=1}^{L}\int_{\Omega_l}(\sigma\nabla\phi)\cdot\nabla\omega d\Omega_l=\sum_{l=1}^{L}\phi_n\int_0^1\sigma\frac{\partial\Psi_n(\xi)}{\partial\xi_k}\frac{\partial\xi_k}{\partial x_j}\frac{\partial\Psi_m(\xi)}{\partial\xi_h}\frac{\partial\xi_h}{\partial x_j}|J(\xi)|d\xi_l \tag{2.61}$$

There is seldom an analytical means of evaluating these integrals so that the numerical integration technique known as quadrature is required. The most efficient quadrature scheme for this integral is Gauss–Legendre quadrature, sometimes referred to as Gaussian quadrature. For a single element, (2.61) describes a total of m × n integrals as there is no implied summation over either of these indices. In a Galerkin formulation m = n and the evaluated integrals can be assembled into an element stiffness matrix, Emn, that is both square and symmetric, resulting in a set of element equations of the form

$$E_{mn}\phi_n=f_m \tag{2.62}$$

Performing the required summation over all of the elements in the solution domain results in a global system of equations that is sparse and also symmetric.

$$\boldsymbol{K\phi}=\boldsymbol{f} \tag{2.63}$$

A zero entry in f is prescribed for all nodes internal to the mesh, representing a conservation of flux at these points. Unlike the FDM, no special treatment of the boundary nodes is needed in order to construct K. On the domain boundary, the same weighting function, ω, can be used to transform the boundary integral from the right-hand side of (2.50) where the dependent variable is a flux of the form ($q = (\sigma\nabla\varphi)\cdot n$).

$$\boldsymbol{f}=\sum_{l=1}^{L}q_n\int_0^1\Psi_n(\xi)\Psi_m(\xi)|J(\xi)|d\xi \tag{2.64}$$

$$\boldsymbol{K\phi}=\boldsymbol{f}=\boldsymbol{Nq} \tag{2.65}$$

After the application of either potential (Dirichlet) or flux (Neumann) boundary conditions at each boundary node, this system can be solved for φ. In practice, the N matrix is rarely constructed and integrated flux values are inserted directly into f.

2.6.4. The Boundary Element Method (BEM)

Like the FEM, the derivation of the boundary element method can begin with the weighted residual form of Laplace' s equation

Also, like the other numerical approximation methods, the endpoint of the derivation can be cast in the form of a matrix as follows. After discretization of its integrals, can be expressed as

$$\phi = g - \mathbf{A}\phi, \tag{2.66}$$

where φ represents the potentials at the nodes of the triangulated boundary surfaces, g are the scaled versions of the same potentials in the virtual, infinite homogeneous medium, and A is the expression of the inhomogeneities. that in an infinite homogeneous medium the secondary sources vanish (σ− = σ+ at all interfaces), and hence, only the first term on the right remains. Formaly, we may write

$$(\mathbf{I} - \mathbf{A})\phi = g, \tag{2.67}$$

from which the desired potential φ can be found as

$$\phi = (\mathbf{I} - \mathbf{A})^{-1} g \tag{2.68}$$

By writing $(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{B}$ we may write

$$\phi = \mathbf{B}g \tag{2.69}$$

and thus express the solution as a simple matrix multiplication. This expression is particularly useful in situations in which the same geometry (and conductivity) apply to a number of sets of potentials, which requires only a change in g as the coefficient matrix, B, remains the same . Until the 1990s, the BEM approach was the dominant method in bioelectric field problems and it still dominates the literature in problems in which current-dipole sources adequately capture the phenomena of interest, most typically in representing discrete sources of bioelectricity in the brain. As we have already seen, in the electrocardiographic forward problem the sources are usually more elaborate, which has led to a wider range of numerical approaches so that today there is a balance between BEM and FEM applications. As already described, the choice of method is multifaceted and depends on both the source formulation and the shape and nature of the volume conductor, and the ultimate goal of the study. Of particular note in the use of the BEM approach are the breakthroughs of Barr et al. in a series of seminal studies leading to

the epicardial potential based electrocardiographic forward (and inverse) problem . Solutions using the BEM approach are typically best suited to conditions in which torso conductivities are at least piecewise constant and isotropic. Of the three numerical approaches, the BEM has the longest history in the field of cardiac bioelectricity.

2.6.5. Geometric Modeling

The three numerical methods already described all assume some form of discrete representation of the geometry of the problem; the creation of these discrete descriptions is what we refer to as geometric modeling. The broader field of geometric modeling contains many additional aspects of computational geometry and computer aided geometry design that seek to capture shape and structure in analytical or statistical form so that its use in the forward problem is just one limited application. In the description to follow, we take a very pragmatic approach and outline the essential requirements for our problem. Geometric modeling is required whenever a simulation problem has a specific geometric context, i.e., it uses neither a schematic representation of actual shape nor a geometrically simplified model made from simple shapes such as lines (planes), circles (spheres), rectangles (cylinders) in two (and three) dimensions. Instead, the geometric models of interest here describe real anatomy derived from some form of images or otherwise digitized reference points and represented as a set of points joined into polygonal elements. Thus a typical geometric modeling pipeline begins with (1) raw geometric data, often a set of medical image data from magnetic resonance (MR) or computed X-ray tomography (CT) from a particular subject or patient. These images sometimes contain specific markers or anatomical fiducial points that provide a reference frame for additional geometry information either from other discrete sources or other image modalities. The next step in the pipeline is (2) to extract from the image data the points (pixels or voxels) that define surfaces between regions of different conductivity. Obviously, the surfaces of interest include the outer boundary of the torso, the surfaces on or inside the heart, and surfaces that surround any other region whose conductivity is considered relevant to the simulation – a determination that is still the topic of research. From these surface boundaries, the next step in geometric model is usually to (3) define a set of suitable points on these boundaries and then connect them into surface polygons (lines in two dimensions and triangles or quadrilaterals in three dimensions). For methods like the BEM, the geometric model is essentially complete at this point but for FDM and FEM, there is an additional step that (4) adds more points and links them into polygons (typically hexahedra or tetrahedra) to describe the volume of the problem domain, e.g., the heart and/or thorax.

2.6.6. Software for Electrocardiographic Forward Problem

From a practical perspective, mathematical, and numerical modeling and simulation approaches are only as good as the software that implements them and that is available for scientists to apply to their own problems. As in almost all research environments, most software created in the laboratories of even the most sophisticated and experienced electrocardiographic researchers is used exclusively in those laboratories and is rarely available for other researchers to use or evaluate. Fortunately, in the area of forward simulations, this situation is better than in many other areas, perhaps a reflection of the relative maturity of at least some segments of this research domain.

Software for Cellular

Software for the simulation of membrane kinetics using the Hodgkin–Huxley formalism described in > Sect. 2.3.1 is relatively widely available, although not always in the flexible, efficient, and simple-to-use form that most serious researchers seek. The simulation approach that is most widely used for cardiac myocytes is probably one of the models from Luo and Rudy and the authors provide a version of the software written in C/C++ at their web site (http://www.case.edu/med/CBRTC/LRdOnline/content.htm). The following is a partial list of software available for download that includes support for cardiac myocyte models

1. Cell electrophysiology simulation environment (CESE): One of the most flexible, portable, and powerful software systems for membrane modeling is a recent projects of Sergey Missan at Dalhousie University, who has created a comprehensive framework that includes implementations for all the leading forms of membrane model .
2. . E-Cell Project: is an international, multicenter research project aiming at developing necessary theoretical supports, technologies, and software platforms to allow precise whole cell simulation that has recently added support for the Luo–Rudy model
3. eb-based program from Semahat Demir et al. at the University of Memphis for carrying out simple simulations of nerve or heart cells. It is very simple to use but there is no access to the source code nor can the application become part of another system.
4. LabHEART: is a model of the myocyte developed by Don Bers and investigators at the Loyola University Physiology Department that includes standard electrophysiological parameters as well as a simulation of calcium concentration in the cell. The application is originally written in LabView but the authors have recently created a new version in MATLAB (

Software for Tissue Simulations

The availability of software for simulation of cardiac tissue is much more limited than for cardiac myocytes, a reflection of the research state of the field and the technical complexity of especially the bidomain forms of simulation. Perhaps the only generally available software for simulation of propagation in cardiac tissue using the bidomain is Cardiowave from Henriquez et al. at Duke University. Cardiowave is capable of creating high-performance simulation programs for a range of platforms, even some that make use of vectorized and parallel computing (http://cardiowave.duke.edu/pmwiki.php). Programming cellular automata is relatively straightforward so that most researchers have developed their own code. There are numerous sources of general purpose cellular automata software.

Software for Volume Conductor Problems

There are many public domain and commercial software packages for the solution of BEM, FEM, and FDM problems but only a small number of dedicated systems have been developed for use in electrocardiographic forward problems. Fortunately, these are some of the best developed and well-supported programs in this application area, providing useful entry points for a range of different types of users.

1. CMISS: CMISS stands for Continuum Mechanics, Image analysis, Signal processing and System Identification and is a massive software system developed by investigators at Auckland University for a broad range of problems, including the electrocardiographic forward problem. The system supports membrane models using the XML markup language known as CellML, bidomain simulations, and volume conductor simulations using both BEM and FEM techniques.
2. ECGSIM: ECGSIM is an interactive program that is very specifically directed at the electrocardiographic forward problem based on the uniform double layer formulation described in > Sect. 2.5.5 of this volume .The user can adjust many parameters of the action potentials and immediately see the resulting changes in all relevant whole heart and body surface parameters (depolarization times, repolarization times, action potential durations and amplitudes, transmembrane potentials, and surface potentials). > Figure 2.13 shows an example of creating a simulation with ECGSIM.

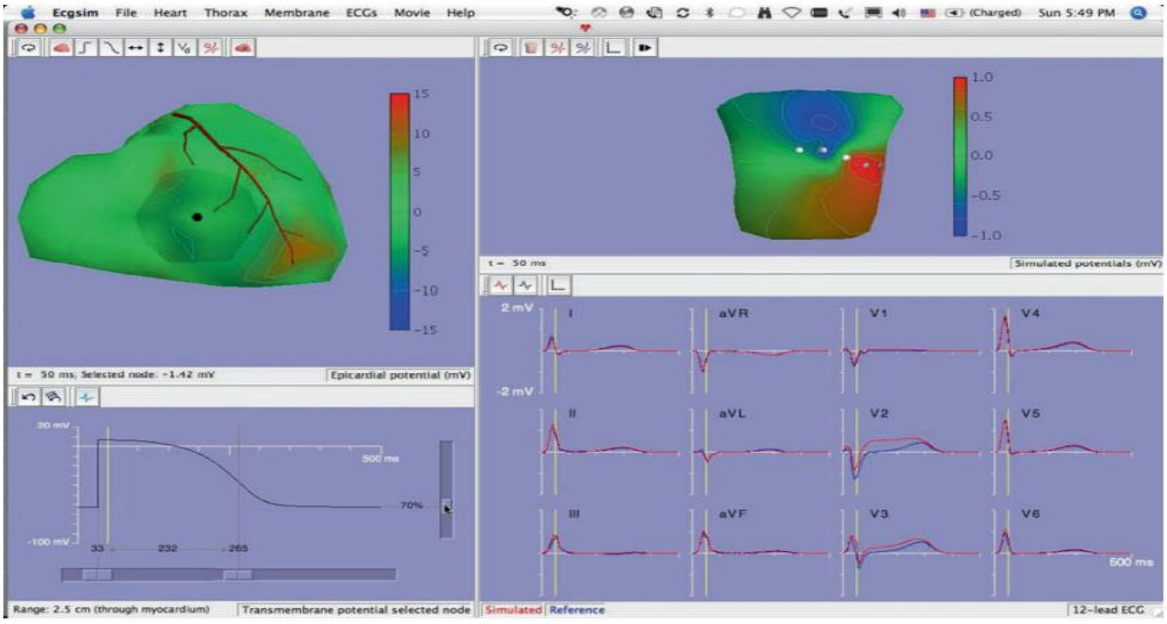

Figure 2-13 User screen for ECGSIM

User screen for ECGSIM. The user can interact with all panels of the screen to adjust display features, select specific time instants, and adjust action potential parameters. The specific example here shows the effect of elevating the resting potential of action potentials in the anterior right ventricle similar to a local transmural ischemic region. Standard ECG signals show the standard shape (in blue) and simulated for this intervention (in red). Yellow vertical lines in the time signals show the time instant displayed in the spatial distributions of potential on epicardium and body surface.

2.6.7. Summary

The forward problem in electrocardiography has become very broad and deep in scope in the last decades as mathematical and computational capabilities allow an unparalleled integration of behaviors in the form of simulations. Progress continues at all levels of the problem as researchers uncover the relationships between the genetic code and ion channel structure and function, examine in more and more detail the linking of cells into myocardium, and use advanced imaging and signal acquisition and processing approaches to capture the shape, structure, and bioelectric fields from the body. At every level of this problem, computational and simulation techniques are essential, not just to gather, analyze, and visualize the experimental data but also to begin to test hypotheses about the roles and interactions of all the elements of this problem. Computers are essential to describe and evaluate possible protein structures and to predict the impact of subtle local changes in action potential shape and duration on the resulting body surface ECGs. The role of the forward problem in cardiac electrocardiography is thus very secure and will provide an intellectual playground for basic questions at all levels of electrophysiology as well as the essential starting point for the inverse problems that we describe in the following chapter.

3. Chapter 3: Forward ECG Modeling

3.1. Introduction

Cardiac electric sources give rise to currents in the body and to potentials in the body and on the body surface. To model the ECG, it is necessary to characterize the bioelectric sources and the volume conductor in which they are immersed.

Thus, a model of the ECG may be considered to consist of: (1) a description of the cardiac sources in terms of the cardiac electrophysiology; and (2) a relationship between these sources and the potentials they generate.

Despite its cellular nature, heart muscle electrically acts in many respects as a syncytium. Furthermore, one can view intracellular space as a continuous domain, or syncytium, with a resistivity that is an appropriate average of the contribution of the ceil interior (cytoplasm) and that of the cell-to-cell connections (tight junctions). A second syncytium is the fluid matrix that forms a continuous extracellular, or intracellular, outer space. The two domains or syncytia are everywhere separated by the cell membrane. By writing the appropriate equations for the currents in the two domains, one can develop an expression that relates the cardiac sources to the cellular transmembrane action potential. We will refer to this approach as the bidomain or bisyncytial model.

The Miller-Geselowitz (MG) model relates the body surface ECG to the distribution of action potentials throughout the myocardium. Much attention has also been devoted to the relationship between the body surface potential distribution and the potential distribution on the heart surface, with particular interest in the inverse problem of determining heart surface potentials from body surface potentials. Thus, there are parallel studies that relate body surface potentials to cardiac sources on the one hand and to heart surface potentials on the other hand. Almost no work has been done, however, to relate heart surface potentials to cardiac sources. Thus, a gap exists in theoretical studies where cardiac sources and heart surface potentials are each related to body surface potentials, but heart surface potentials have not been related to cardiac sources.

3.2. Theory

First, we note that the current density, J, in regions beyond the site of an action potential is linearly related to the electric field intensity, E. Furthermore, the capacitive component of tissue impedance is negligible at frequencies below several kilo Hz of interest to electrocardiography, and in addition there is evidence that pulses with rise times of the order of a microsecond suffer only minor distortion through the thorax

Therefore, to a good approximation a region containing no bioelectric sources can be assigned a uniform bulk conductivity σ so that

$$J = \sigma E \tag{3.1}$$

A second important point is that electromagnetic wave effects can be neglected. Hence, at each instant the electric field can be obtained from the gradient of an electric scalar potential, ϕ.

$$E = -\nabla\phi \tag{3.2}$$

As a consequence of these properties of body tissues, the currents at any instant depend only on the sources at that instant and obey the principle of superposition. Formally we can represent the sources by a distribution of impressed current density or equivalently, by current dipole moment per unit volume, J^i. According to this formalism, equation (1) is modified to include active regions as follows, σE is the conduction current, J^i current density source:

$$J = \sigma E + J^i \tag{3.3}$$

Later we shall attempt to relate J^i to electrical activity associated with the plasma membranes of the active cells. For the present, however, the sources are taken into account by the addition of impressed currents to the appropriate conducting regions. Neglect of tissue capacitance implies that as the sources vary, charges on boundaries and interfaces redistribute themselves in a negligibly short time, or that equivalently

$$\nabla . J = \frac{\partial \rho}{\partial t} \tag{3.4a}$$

$$\nabla . J = 0 \tag{3.4b}$$

which can be combined with equations (2) and (3) to give

$$\nabla . \sigma \nabla \phi = \nabla . J^i \tag{3.5}$$

Let the surface Sj separate regions of conductivity $\sigma'$and σ", and let dS_j be a differential element of area of this surface. Adopt the convention that dSj, is directed from the primed region to the double primed one. Since the current must be continuous across each boundary,

$$\sigma' \nabla \phi' . dS_j = \sigma'' \nabla \phi'' . dS_j \tag{3.6}$$

Furthermore, the potential is also continuous at each boundary. Hence,

$$\phi'(S_j) = \phi''(S_j) \tag{3.7}$$

Let dv be an element of volume of a homogeneous region, v, and let A and B be two functions which are well behaved in each region. Green's theorem then states that

$$\int_s (A\nabla B - B\nabla A).\bar{n}dS = \int_V (A\nabla^2 B - B\nabla^2 A).dv \tag{3.8}$$

With

$$A = \frac{1}{\vec{R}} = \frac{1}{r-r_o} \;\; and \;\; B = \sigma\phi \tag{3.9}$$

As r is the location of the source and r_o is the location of the observation point

$$\int_V \sigma\phi\nabla^2 \frac{1}{\vec{R}}.dv \;=\; -4\pi\sigma\phi(r_o) \tag{3.10}$$

as $\lim_{r=r'} \frac{1}{\vec{R}} \;=\; \lim_{r=r'} \frac{1}{r-r'} \;=\; -4\pi\delta(|r-r'|)$

Then:

$$\sum_j \int_{s_j} [\frac{1}{\vec{R}}(\sigma'\nabla\phi' - \sigma'\nabla\phi'') - (\sigma'\phi' - \sigma''\phi'')\nabla\frac{1}{\vec{R}}].\bar{n}dS \;=\; \int_V \frac{1}{\vec{R}}\nabla.\sigma\nabla\phi dv + 4\pi\sigma\phi(r_o) \tag{3.11}$$

where $\sigma \; and \; \phi$ in the last term are evaluated at $\vec{R}$ = 0, i.e., the observation point.

$$4\pi\sigma\phi(r_o) = -\int_V \frac{1}{\vec{R}}\nabla.J^i dv - \sum_j \int_{s_j} \phi(\sigma' - \sigma'')\nabla\frac{1}{\vec{R}}.\bar{n}dS \tag{3.12}$$

The first integral on the right may be transformed using the divergence theorem as follows:

$$\int_V \nabla.(J^i/R)\, dv = \int_V J^i/R\, ds = \int_V (J^i.\nabla\frac{1}{\vec{R}} + \frac{1}{\vec{R}}\nabla.J^i)\, dv \tag{3.13}$$

If J^i vanishes on S, the boundary of the region containing the sources, then:

$$\int_V \frac{1}{\vec{R}}\nabla.J^i\, dv = -\int_V J^i.\nabla\frac{1}{\vec{R}}\, dv \tag{3.14}$$

Substituting in (3.12) we get:

$$4\pi\sigma\phi(r_o) = \int_V J^i.\nabla\frac{1}{\vec{R}}dv - \sum_j \int_{s_j} \phi(\sigma' - \sigma'')\nabla\frac{1}{\vec{R}}.\bar{n}dS \tag{3.15}$$

For the volume conductor the right summation of surface integrals contains only one surface, which is the surface of the body, and the current density

$J^i = 0$ everywhere except inside the, then the potential at the r_o on the surface of the body:

$$\phi(r_o) = \frac{1}{4\pi\sigma}\int_{V_H} J^i . \nabla \frac{1}{\bar{R}} dv - \frac{1}{4\pi}\oint_{S_B} \phi_B \nabla \frac{1}{\bar{R}} . \bar{n} dS \tag{3.16}$$

Using the bidomain method, We can relate current density J^i to the transmembrane potential ϕ_m, as σ_i is the conductivity inside the heart:

$$J^i = -\sigma_i \nabla \phi_m \tag{3.17}$$

Then:

$$\phi(r_o) = -\frac{\sigma_i}{4\pi\sigma}\int_{V_H} \nabla\phi_m . \nabla \frac{1}{\bar{R}} dv - \frac{1}{4\pi}\oint_{S_B} \phi_B \nabla \frac{1}{\bar{R}} . \bar{n} dS \tag{3.18}$$

Using Green's first identity:

$$\int_D \nabla U \cdot \nabla V dV + \int_D U \nabla^2 V dV = \int_{\partial D} U \nabla V \cdot n dS \tag{3.19}$$

And given that $\nabla^2 \phi_m = 0$, we get:

$$\phi(r_o) = -\frac{1}{4\pi\sigma}\int_{S_H} \sigma_i \phi_m d\Omega_H - \frac{1}{4\pi}\oint_{S_B} \phi_B d\Omega_B \tag{3.20}$$

Given that: $d\Omega = \nabla \frac{1}{\bar{R}} . \bar{n} dS$. (3.21)

By discretizing the surface integrals, and substituting at every point on the body surface:

$$\phi_k = \frac{-\sigma_i}{4\pi\sigma}\sum_{j=1}^{N} \phi_{m,j}\Omega_{Hjk} - \frac{1}{4\pi}\sum_{j=1}^{M} \phi_{Bj}\Omega_{Bjk} \tag{3.22}$$

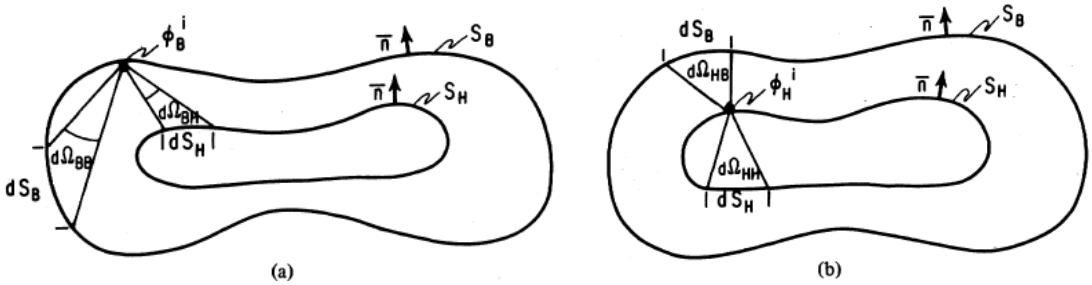

Figure 3-1 (a) An observer on outer surface S_B. And (b) An observe on inner surface S_H. Vector $\bar{r}$ extends from the observation point to the element of integration dS_f.

This equation when applied to a point on the torso, will give us one linear equation containing all M torso elements potentials and all N heart elements potentials. By substituting in this equation at each torso point, it will give us a system of linear equations containing M equations and M variables (body surface potentials). Solving this linear system of equations will give us the torso potentials knowing the heart transmembrane potentials Φ_H.

From the discretization, we get the value of each element in the matrices $P_{BB}\ and\ P_{BH}$.

$$P_{BB}^{jk} = \begin{cases} 1 + \frac{1}{4\pi}\Omega_{Bjk} & , i = j \\ \frac{1}{4\pi}\Omega_{Bjk} & , i \neq j \end{cases} \tag{3.23}$$

$$p_{BH}^{ij} = \frac{-\sigma_i}{4\pi\sigma}\Omega_{Hjk} \tag{3.24}$$

$$P_{BB}\Phi_B = P_{BH}\Phi_H \tag{3.25}$$

$$\Phi_B = P_{BB}^{-1} * P_{BH}\Phi_H \tag{3.26}$$

$$\Phi_B = Z_{BH}\Phi_H \tag{3.27}$$

Thus: $$Z_{BH} = P_{BB}^{-1} * P_{BH} \tag{3.28}$$

The above technique is the boundary element method.

Alternatively, one can solve the problem using finite element techniques. Pilkington, Morrow, and Stanley have provided comparative data on the two approaches. The finite element technique would probably be the method of choice if anisotropy is to be incorporated in the model.

3.3. Bidomain derivation

The essence of the relationship between dipole sources and cellular activity was developed by Wilson, MacLeod, and Barker. Considering a single cylindrical fiber, they showed that the dipole moment was related to the spatial gradient of transmembrane action potential. This concept was confirmed in tissue bath experiments in one and two dimensions by Spach and co-workers. Following is a theoretical development which generalizes this concept to three dimensions by considering heart muscle to be composed of two domains.

Cardiac muscle cells are cylindrical, tend to be arranged in parallel arrays in local regions, and are interconnected in a complex fashion. Connections between cells exhibit a low resistance in the normal state. Therefore, despite its cellular nature, heart muscle electrically acts in many respects as a syncytium. One can view intracellular space as a continuous domain, or syncytium, with a resistivity which is an appropriate average of the contribution of the cell interior (cytoplasm) and that of the cell-to-cell connections (tight junctions). A second domain is the fluid matrix which forms a continuous extracellular, or interstitial, outer space.

Following a suggestion of Schmitt, Miller and Geselowitz approached the representation of sources in the heart by considering the heart to consist of an inner and an outer compartment, or domain, each of which is a passive conductor. The two domains are everywhere separated by the cell membrane. We consider the two domains to be each smeared out over the volume of the heart muscle, as is the membrane.

The model may be considered a macroscopic one in that current densities, potentials, and conductivities are averages over regions containing many cells. At the microscopic level extracellular potentials close to the cells reflect the cellular nature of the tissue, but this effect becomes smaller at larger distances.

If J is the current density, Φ is the potential, and σ is the conductivity, then

$$J_i = -\sigma_i \nabla \phi_i \tag{3.29a}$$

$$J_o = -\sigma_o \nabla \phi_o \tag{3.29b}$$

where the subscripts i and o represent intracellular and outer (extracellular) space respectively.

Charge moving from one domain must cross the membrane and enter the other domain, giving rise to a membrane current per unit volume I_m. As noted, current is conserved in the absence of an external source. The mathematical statement of these facts is

$$I_m = -\nabla . J_o = \nabla . J_i = -\nabla . \sigma_i \nabla \phi_i \tag{3.30}$$

where the positive direction is taken to be outward. By definition, the transmembrane potential ϕ_m, is

$$\phi_m = \phi_i - \phi_o \tag{3.31}$$

Combining these equations leads to the following relationships:

$$J = J_i + J_o = -\sigma_i \nabla \phi_i - \sigma_o \nabla \phi_o = -\sigma_i \nabla \phi_i - \sigma_o \nabla \phi_o + \sigma_i \nabla \phi_o - \sigma_i \nabla \phi_o$$

$$J = -\sigma_i (\nabla \phi_i - \nabla \phi_o) - (\sigma_o + \sigma_i) \nabla \phi_o$$

$$J = -\sigma_i \nabla \phi_m - \sigma_H \nabla \phi_o \tag{3.32}$$

$$\nabla . J = 0 = -\nabla . \sigma_i \nabla \phi_m - \nabla . \sigma_H \nabla \phi_o \tag{3.33}$$

$$\nabla . \sigma_i \nabla \phi_m = -\nabla . \sigma_H \nabla \phi_o \tag{3.34}$$

where $\sigma_H = \sigma_i + \sigma_o$, is the bulk conductivity of heart muscle. The bidomain model then gives the following equations for the potential in the heart.

Formally we can represent the sources by impressed currents. If j^i is the impressed current density at a point, then the total current J is given by

$$J = J^i - \sigma \nabla \phi \tag{3.35}$$

Since the divergence of J vanishes,

$$\nabla . J^i = \nabla . \sigma \nabla \phi \tag{3.36}$$

A comparison of Equations. (3.34) and (3.34) indicates that the current source distribution, J^i, is given by

$$J^i = \sigma_H \nabla \phi_o \tag{3.35}$$

$$J^i = -\sigma_i \nabla \phi_m \tag{3.36}$$

3.4. Modelling of The Heart

One way the simulation works is that it checks for each individual node' s depolarization time and the current time step and depending on the supposed depolarization time for a cell compared to the current time step, the cell either depolarizes or stays as is.

Suppose cell A, represented as a node on the simulation, is depolarized at time $t = 5\ ms$, And its neighboring cell B has a delay time of $10\ ms$, and the time step is $1\ ms$. The simulation will check on cell B every time step as per the condition that

$$t_B > t_A + t_{delay}. \tag{3.37}$$

This concept is called Cellular Automata. A cellular automaton is where the next state of cell is affected directly by its current state and the state of its neighboring cells. In other words, here we see that the state of cell B at time t can be represented as a function of the state of cell A at time $t + t_{delay}$, as mentioned in (3.35).

Applying the same concept on a bigger scale affecting all the cells of the heart, the heart depolarization wave can be represented. The cellular automata concept was mentioned in Sect 2.4.2. The following figure explains how the wave propagates depending on the principles of cellular automata rules.

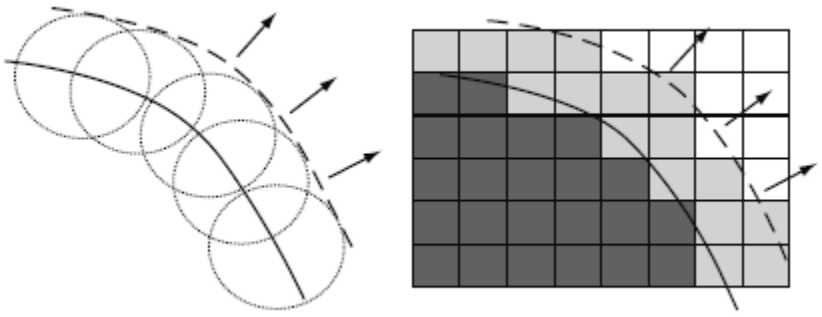

Figure 3-2 Wave propagation applying cellular automata

Using a constant wave speed approach. The new position of the wavefront can be found by placing circles of an appropriate diameter on the existing wavefront (left). With previously excited cells drawn dark grey and quiescent cells white, the cells which

The modelling of the heart for the simulation is the Equivalent Double Layer source model. This model has the potential of describing the cardiac electric generator during the depolarization phase 'The QRS segment' as well as the during the repolarization phase of the cardiac cycle "The STT segment".

The EDL source is an equivalent expression for uniform double-layered sources at the propagation wave fronts during depolarization. The effect of repolarization, which starts right from the moment of the first local depolarization, is disregarded in UDL sources. Thus, after depolarization is complete, there's no external potential field. This is the moment at the end of the QRS complex in the ECG, the J point.
A more complete model of the sources of the ECG would therefore require the inclusion of the repolarization. The time interval following the J point, the STT segment, has been long recognized as yielding highly relevant diagnostic information. The generalization of the EDL source provides a model valuable for describing the current sources of the ECG throughout the entire cardiac cycle.

Anatomically, the heart consists of three main layers; The epicardium, which is the outer layer of the heart wall. The endocardium, which is the thin inner layer of the heart wall. The myocardium, which is the middle layer between the epicardial layer and the endocardial layer. This is taken into consideration when modelling the heart as an EDL source. The following figure helps explain the idea of the EDL source further.

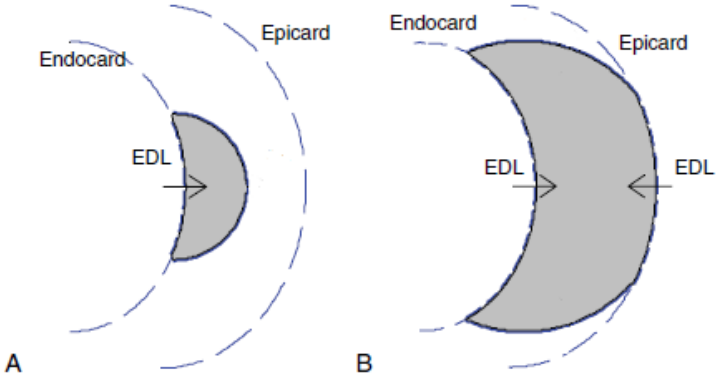

Figure 3-3 Depolarization on EDL surfaces

The Shaded region represents depolarized myocardium; arrows denote polarity of EDL double strength. Dotted lines represent EDL locations. Panel A*: Early stage of depolarization following endocardial stimulus.* Panel B*: later stage of depolarization; wave front previously broken through at the epicardium.*

The heart can be cut down into a number of 2d triangles in order to be represented in a 3d view. The potential of a point on the torso is calculated thru the equivalence of potentials of the corresponding

triangle. Consider a single triangle, the potential of the corresponding point on the torso is calculated through the process explained in Fig 3.1. Note that in 3d modelling, the solid angel of integration is conically-shaped and can be calculated from the equation (3.21). When applying a closed surface integration on the heart as a whole, depolarization occurs firstly on a point on the endocardial layer. its potential increases and its contribution' s direction is that of the vector to the observer. Following this, a point on the epicardial layer is depolarized. Its potential increases but with a time delay from the first point. And the direction of its contribution is the opposite of direction of the vector to the observer.

3.5. Simulation of The Heart

When calculating the heart and torso potentials by using machine learning, a large number of nodes is needed. Having the number of nodes over the torso be larger than the number of nodes over the heart is a must in order to calculate the potential matrices. When using machine learning, we apply probes on the heart and other probes on the torso. Getting the values of parameters of such probes for each time step, it' s noticeable that the values differ for each time step.
When studying a specific set of probes on the torso and another set of probes on the heart and exporting the torso set' s values to the machine learning software and after the machine learning predictions and calculations, we get the expected values of the probes of the heart ready to be presented.

The machine learning module receives the values of the probes on the torso, in the form of Φ_B ,and return the values of the probes on the heart, in the form of Φ_H. Still, an issue occurs where the number of probes on the heart is way less than the number of nodes on it. That' s where an interpolation of the nodes takes place with respect to the values of the probes and the distances between them and surrounding nodes.

By establishing and running a simulation of the heart and its probes and potentials, utilizing the forward problem this way, it is possible to solve the inverse problem via the machine learning module, which utilizes the data received from the forward simulation of the heart.

The following figure represents the heart modelling during depolarization and repolarization. From left to left, the atria depolarize, then the ventricular depolarization occurs, and finally the repolarization starts in the ventricle.

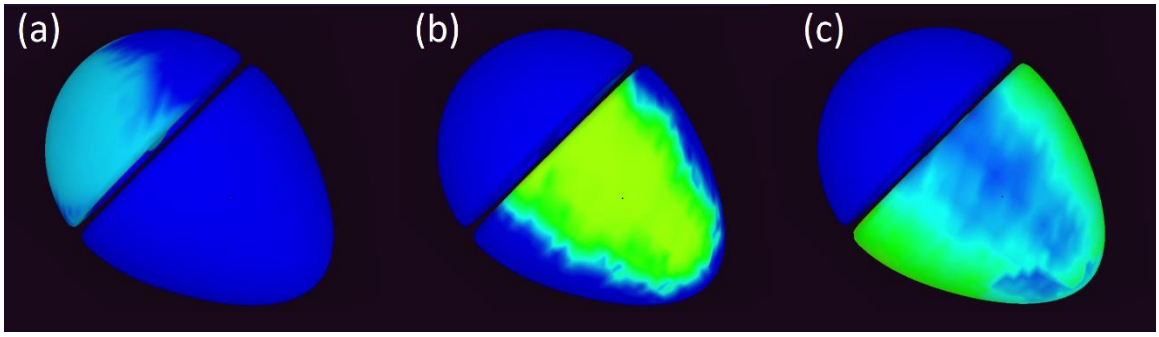

Figure 3-4 The effect of the polarization cycle on the potentials of the torso

The Polarization phases of the heart starts with (a) the depolarization of the atria where the SA node pumps into the AV node, which represents the P-wave of the ECG and is depicted by the left model. Then (b) the ventricular depolarization where the AV node pushes the blood through the His bundle to the left-side and right ventricles, which is the QRS-wave in the ECG and is depicted by the middle model. And finally, (c) the repolarization phase where ventricles pump the blood to the rest of the body, which is the T-wave in the ECG and is depicted by the right-side model.

The effect of this can be shown on the torso. Representing the potentials of the torso in the Fig 3.5.

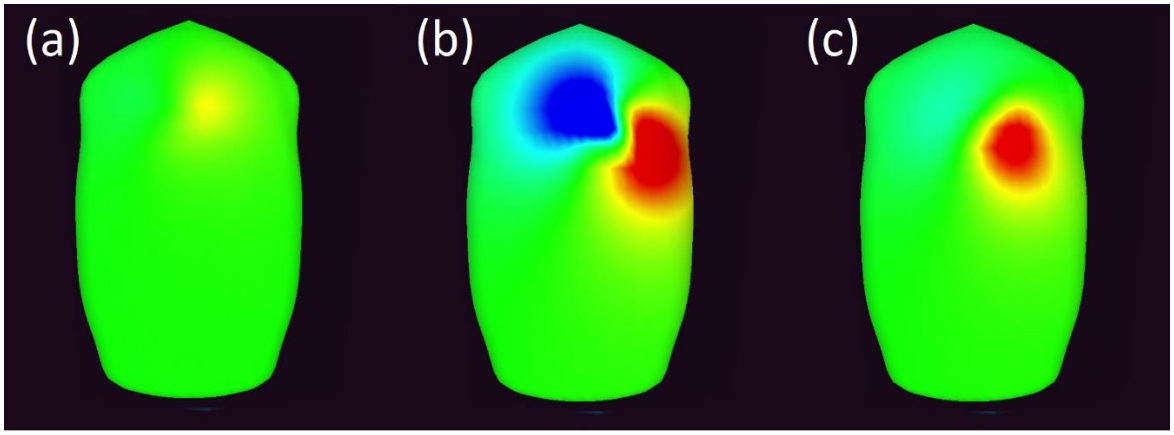

Figure 3-5 The effect of the polarization cycle on the potentials of the torso

At (a) and (b) the effect of depolarization takes place, also affecting the nodes V1 to V6 and the left, right, and foot leads. Where (a) is the depolarization of the atria and (b) is the depolarization of the ventricles. (c) shows the effect of the repolarization on the torso.

As discussed in the first chapter, The ECG signal consists of 3 main phases: The P-wave, the QRS-wave, and the T-wave. And such phases are monitored by checking the ECG diagrams of the nodes V1 to V6 and the left, right, and leg leads aVL, aVR, and AVF respectively.

Checking the Fig 3.6, we can see that for some nodes, waves like the P-wave and the QRS-wave are close to the default. Whereas they might differ in one way or another for other nodes. The aspects at which the shape and value of these waves differ are the strength and direction of mention waves.

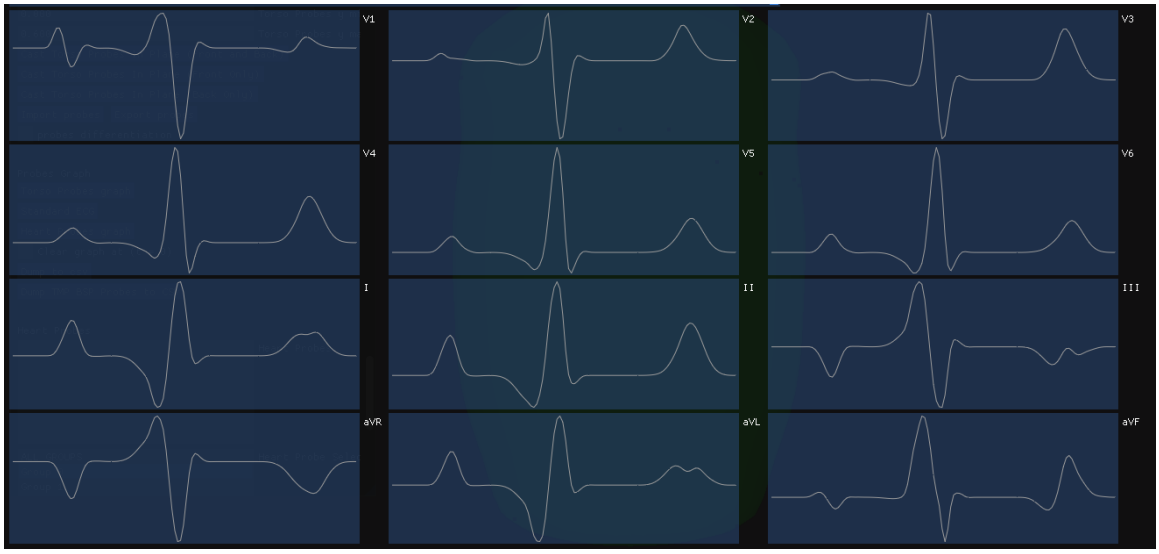

Figure 3-6 The ECG waves of the main leads and nodes of the heart

4. Chapter 4: The Inverse Problem of Electrocardiogram

4.1. Introduction

In the inverse problem one would like to invert the forward model, or equivalently solve for the source parameters which, together with the forward model, explain or predict the measurements. In contrast to the forward problem, which can be solved uniquely to within a constant for the zero of potential, the inverse problem does not possess a complete solution that is mathematically unique. By this we mean that it is not possible to identify unique cardiac sources within the heart as long as the active region is inaccessible, because the electric field that they generate outside any closed surface enclosing them may be duplicated by equivalent single- or double-layer sources on the closed surface itself. This difficulty is usually circumvented by using simplified models for the cardiac sources; in this case, the parameters of these models may be uniquely determined from the surface potentials with additional assumptions regarding the intervening volume conductor.

It is not yet clear what the limits are on the degree to which one can image intramural activity from body surface or intracavitary measurements, nor how reliable such imaging may be. We discuss in some detail in the following section how to use temporal behavior to improve inverse solutions for surface models, as well as some recent, initial attempts to reconstruct intramural activity.) A simple description of a parametric model for the inverse problem is the vectorcardiogram (VCG), which assumes that the heart' s electrical activity may be approximated by a fixed-location, variable-amplitude, variable-orientation current dipole within a finite homogeneous torso. More sophisticated inverse solutions upgrade the heart source model by using a more realistic description of cardiac depolarization /repolarization. Along with these upgraded heart models, most contemporary simulations include a realistically shaped and often patient-specific geometric model of the torso. Early formulations of the inverse problem in electrocardiography treated it as a kind of extension of the traditional electrocardiogram (ECG), based on measurements made on the body surface. With the advent of intracavitary catheter-based probes for clinical electrophysiology (EP) procedures, researchers also began to use measurements from multielectrode noncontact probes, located inside one of the heart chambers, in an inverse solution to reconstruct the electric potential on the inner wall of the chamber. In the literature, this has been called the "endocardial inverse problem."

we concentrate on source models that capture parameters that one can physically measure, specifically surface potentials (epicardial and/or endocardial), transmembrane potentials on the epicardial and/or endocardial surfaces. Traditionally, potential-type sources have used a single surface model (the epicardium for body-surface measurements, the endocardium for intracavitary measurements), while fiducial-time based approaches have by necessity used a combined endocardial/epicardial surface. Since any surface which encloses all the sources is a valid surface on which to model the sources, potential-based methods can also use a combined surface.

It is impossible to recreate the electrical state of each cell in the heart (or even each small cluster of cells) from surface or intracavitary electrical recordings, no matter how many recordings are available. However, the impossibility of this reconstruction is a mathematical certainty – perhaps most simply explained by the fact that multiple configurations of cellular activity can give rise to the same measurements (that is to say, the problem is not unique). If, however, attention is restricted to finding some alternative parameters to represent the electrical activity of the heart, then it is possible to construct an inverse problem with a mathematically unique solution. Unless this problem is appropriately constrained, the solution will not depend continuously on the data, meaning that small perturbations in the input data will result in disproportionately large changes in the computed solution. This is illustrated in Fig. 4.1c, which shows computed epicardial potentials that best match the measured body surface potentials in a least-squares sense; with no constraints the results are completely erroneous. An important consequence of nonuniqueness and ill-posedness is that the level of detail that one can reconstruct from body surface electrical recordings is significantly lower than that which can be simulated with a similarly resolved forward solution.

We can consider the difficulty and the effects of ill-posedness in an electrocardiographic inverse problem as an imaging problem. The goal in this case is to reconstruct the electrical state of the heart just as anatomy is reconstructed using an inverse solution in the more widely known imaging modalities of magnetic resonance imaging (MRI) or computed tomography (CT). In MRI, the forward model is a (discrete) Fourier transform, and thus, there is no difficulty inverting it reliably to create an image. However, for the inverse ECG problem, the forward solution includes attenuation that is of different degrees for different spatial scales. In particular, higher spatial frequencies (smaller-scale phenomena) are attenuated more than those with lower frequencies. As a consequence, trying to invert the forward solution directly leads to the amplification of high-frequency components. Of course, any noise in the measurements and any error in the forward model will contain such high-frequency components which will then dominate unconstrained solutions, as seen in Fig. 4.1; the effects of these components on the solution must be constrained.

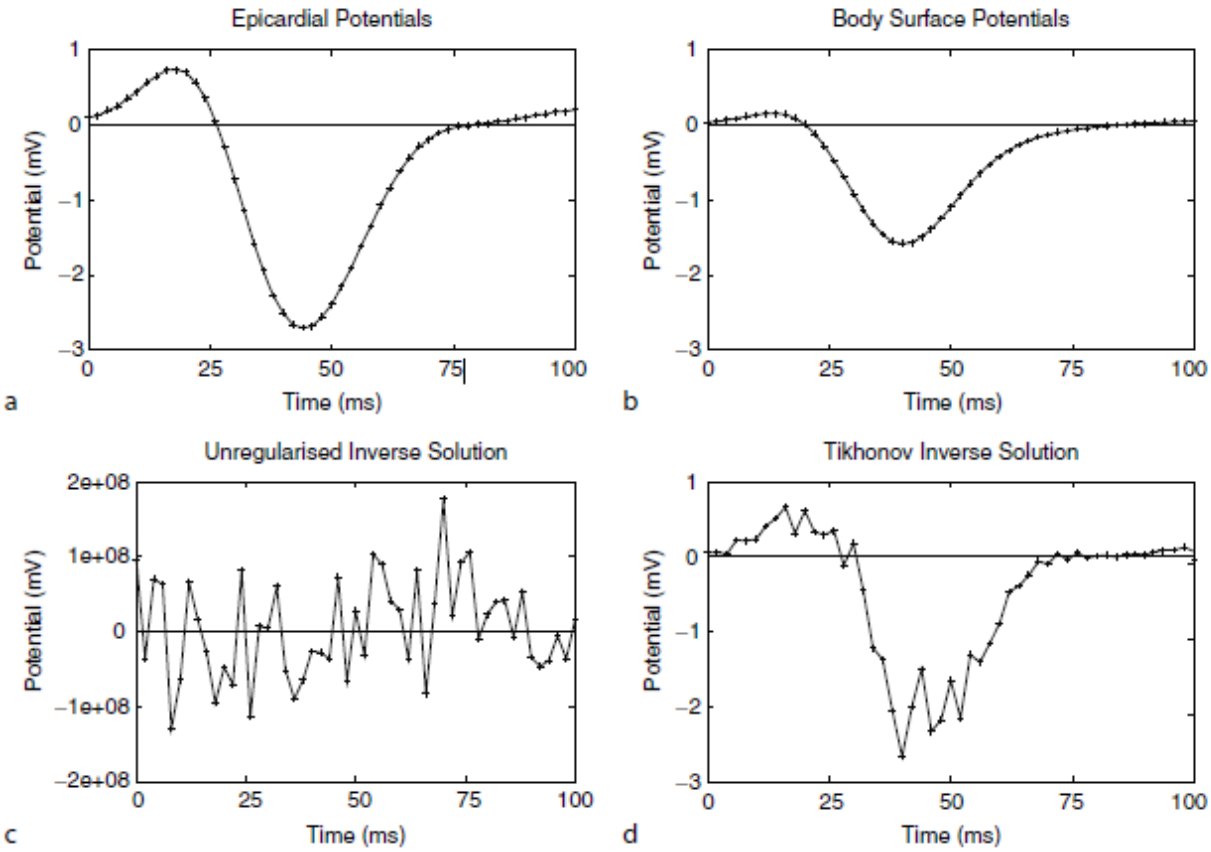

Figure 4-1 ***An illustration of ill-posedness***

Shown are (a) epicardial and (b) body surface signals generated in a forward simulation. Also shown are resultant inverse solutions for (c) an unconstrained inverse solution taken from the data in (b); and (d) a Tikhonov regularized inverse also using data in (b). The unconstrained epicardial signals are clearly erroneous, highlighting the ill-posedness of such an inverse problem.

4.2. Inverse Problem Formulation

In this section we describe in general terms a formulation of the inverse problem that encompasses all the variants we will then describe in more detail. We will then discuss, again in general terms, solution strategies that address the ill-posedness explained in the previous section. Two requirements for solving an electrocardiographic inverse problem are: (1) a mathematical (geometric) description of the region through which the electrical currents generated within the heart flow to the recording sensors, whether they are on the torso surface or in the intracavitary blood volume and (2) recordings of those electrical signals themselves, sampled at tens to perhaps two hundred known locations, at typically between 500 and 2,000 samples per second.

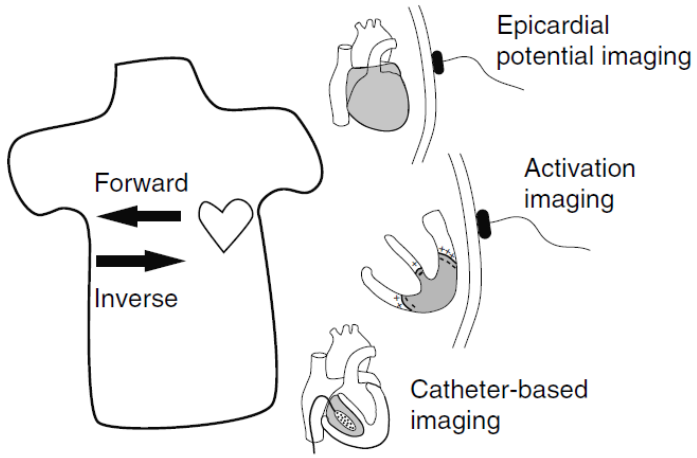

Figure 4-2 the various types of inverse solutions

There are three basic types of inverse problems treated in this chapter, as illustrated here: in two of them, body surface measurements are used to reconstruct potentials (surface or transmembrane) or activation wavefront locations (isochrones), and in the third intracavitary potentials are used to reconstruct surface potentials. The previous figure illustrates the various types of inverse solutions we consider here. Measurements of body surface potentials can be used in one of two ways depending on the source and forward model to which they are applied: either to reconstruct potentials (surface or transmembrane) or to reconstruct fiducial times such as timing of wavefront arrivals on the epicardial and endocardial surfaces (activation times). Measurements of intracavitary potentials can be used to reconstruct surface or even transmembrane potentials as well, although to date this technique has usually been applied to reconstruct endocardial potentials.

To formalize this setting, we first summarize in a general form the results of a solution to the forward problem.

$$b = A(x) + n, \tag{4.1}$$

where A is a (possibly nonlinear) forward operator which incorporates a parameterized model of the cardiac sources and produces a prediction of the measured potentials, x holds the solution of interest (e.g., epicardial potentials or activation isochrones), b holds the measurements (whether body-surface or intracavitary), and n represents the measurement noise. For linear forward operators, A becomes a matrix A and measurements, solution and noise can be denoted as column vectors: b, x, and n respectively. In this situation the forward model becomes

$$b = Ax + n, \tag{4.2}$$

Although we have written the forward operator A above as free of error, it is important to recognize that generally none of the components of (4.1) or (4.2) are exact – in addition to measurement noise the forward

operator is corrupted to some degree by the geometric noise arising from many sources: inaccuracies in developing the geometric model, errors associated with the numerical solution of the forward problem, errors in conductivities, etc. Moreover, the measured signals have errors in the recording of the positions of the electrodes. Errors may be considered small (e.g., signal noise below 10 microvolts, submillimeter MRI or CT image accuracy) but such errors can and do become greatly amplified in an inverse calculation, especially without careful treatment, due to the ill-posed nature of the problem.

4.2.1. Including Time in the Formulation

There are a number of ways that temporal aspects of the inverse problem can be formulated. We can consider x, b, and n as representing an entire temporal sequence of the relevant parameters in the case of potential sources. Alternatively, b and n could contain temporal sequences, while x could be the relevant fiducial timings for activation or recovery time source models. An alternative formulation for potential models could be to consider x, b, and n as spatial samples at a given time instant, and then write a temporal sequence of equations of the form of (4.1), one for each sample instant.

In the context of a linear model, x, b, and n in (4.2) would be block vectors formed by concatenations of the different time instants, and the matrix A would then be a large block diagonal matrix, with the forward matrix repeated in each of the diagonal blocks. Written explicitly, then, the vector x becomes

$$x = [x(t_0)^T x(t_0 + 1)^T \dots x(n)^T]^T, \tag{4.3}$$

where $x(t_i)$ is the vector formed by stacking the potentials over the solution surface at time t_i and the time instants taken into consideration range from t_0 to t_n. The block vectors for b and n would be written similarly, and then A would have the form (where here we use $\widetilde{A}$ to represent the single time-instant forward matrix):

$$A = \begin{bmatrix} \widetilde{A} & 0 & \cdots & 0 \\ 0 & \widetilde{A} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \widetilde{A} \end{bmatrix}, \tag{4.4}$$

Another useful formulation is to simply stack each time sample as one column of a matrix and then re-write the forward solution over the entire time sequence as

$$B = AX + N, \tag{4.5}$$

where the matrices B, X, and N have columns b, x, and n, which are spatial vectors at discrete locations, and each row is a time series at a given spatial position.

Note that there is no explicit time dependence in the forward operator A in any of the above equations. Indeed, in the inverse problem, A is generally taken to be temporally invariant, defined using the geometry of the heart in diastasis without consideration of the dynamic variation of that geometry. One could

certainly argue that ignoring the dynamic nature of A (which arises as a result of heart motion, among other sources of physiological motion) is a possibly meaningful source of error. However, there are a number of justifications that support this static assumption. First, because of the delay between excitation and contraction in cardiac myocytes, significant contraction does not occur until after the QRS complex. As a result, ignoring contraction should be a reasonable approximation for imaging the depolarization of the heart. This argument is, of course, not valid for imaging repolarization, and the importance of this additional error source in repolarization imaging has yet to be determined. However, perhaps the strongest reasons for treating A as a static matrix are pragmatic ones. It is already difficult enough to obtain and use image data (e.g., MRI or CT) to create a mesh of the heart and surrounding torso volume of a specific patient at one time instance; repeating this process to construct multiple transfer operators would be prohibitive, given the present state of model-creation technology. One recent attempt to mitigate this problem has been the suggestion to use electrical impedance tomography, which may allow one to image changes in the geometry.

4.2.2. The Inverse Problem

The inverse problem can be stated very succinctly as finding a solution x which (1) matches the measurements that the model A would predict for that particular solution to the actual measurements b and (2) also is indeed a reasonable solution for parameters of cardiac electrical activity. One can think of this as "inverting" the operator in (4.1). However, there are iterative approaches to finding x that do not explicitly compute an inverse, for either algorithmic or computational reasons.

Even with a linear forward matrix and x representing a vector of epicardial potentials ϕ, the inverse problem is ill-posed. There is no reason to expect A to be square, and therefore, it is typically not invertible. The usual solution in such cases, at least if the number of measurements is greater than the number of potentials to be solved for on the solution surface (A is over-determined), is to solve for the least-squares solution. This is the solution which minimizes the (Euclidean) magnitude of the residual error $\|A\phi - \phi_B\|_2$ (where ϕ represents the solution x for the specific case of heart surface potentials, and ϕ_B represents b for the specific case of body surface measurements). The resulting solution solves the square matrix equation

$$\phi = (A^T A)^{-1} A^T \phi_B \tag{4.6}$$

This equation has a unique solution if the columns of A are linearly independent. However, even in this case, $A^T A$ is even more poorly conditioned than A, meaning that its inverse is even more challenging to

obtain reliably. If the number of measurements is less than the number of cardiac locations at which one desires to estimate the relevant parameters, (i.e., the matrix A is under-determined) there can be no unique solution. The most common strategy in this case is to find a solution that minimizes the Euclidean magnitude of the residual and is itself the shortest (in Euclidean length) of the infinite number of vectors that will achieve this minimum error. The formula for finding the solution is similar in spirit to (4.6). It too requires solving a very badly conditioned system of equations, which leads to the same problem as discussed in the framework of (4.6). Thus, since the forward operator is ill-posed, a simple data fit by the minimization of the residual norm $\|A(x) - B\|_2^2$, for any of these formulations, leads to a wildly erroneous inverse solution, with unrealistically large magnitudes, as illustrated in Fig. 4.1. To obtain a reasonable inverse solution, prior knowledge about the solution needs to be added to the problem formulation as a constraint or set of constraints. The difficulties in accomplishing this goal include identification of physiologically useful descriptions of such prior knowledge, of mathematically tractable ways of including them into the inverse problem, and of practical algorithmic approaches to solving the resulting constrained optimization problem.

The problem of determining $\boldsymbol{x}$ in (4.2) is an example of a rather general type of "inverse problem" that arises frequently in science and engineering.

In inverse electrocardiography, common formulations fall into two categories. One is a deterministic framework, generally referred to as "regularization," in which an objective function to be minimized or a constraint function to be satisfied is composed of a combination of the norm of the residual error and some norm of a constraint function. The other category of formulation is a statistical framework, in which the solution is treated as random with an appropriate probability model, and a probabilistic error measure is minimized to find a likely solution. Conceptually, deterministic regularization approaches can be summarized by two different ideas:

1. *Approximate the best-fit-to-data solution.* One or more spatial penalty functions acting on a candidate solution are defined, and then a weighted sum of these functions plus the residual norm is minimized. This approach is generally called Tikhonov regularization. The penalty functions generally used constrain the magnitude of the inverse solution or its high spatial frequency content (often formulated via a first or second order spatial derivative). One can summarize the resulting formulation as:

$$\hat{x}_\lambda = argmin\{\|A(x) - b\|_2^2 + \lambda^2 \|R(x)\|\}, \tag{4.7}$$

where R represents a "regularization operator" and λ is the regularization parameter whose value controls the level of regularization (i.e., the balance between the data fit and the amount of regularization). For a nonlinear operator A or nonlinear regularization operator R, nonlinear optimization methods are used to obtain the solution. For the linear cases, A and R become matrices and the Tikhonov solution simplifies to:

$$\hat{x}_\lambda = (A^T A + \lambda^2 R^T R)^{-1} A^T b \tag{4.8}$$

(or equivalently solving $(A^TA + \lambda^2 R^T R)x_\lambda = A^T b$ by any appropriate algorithm).

2. *Approximate the forward operator.* In this approach, the forward operator A is approximated with an operator A_r that is "similar" to A in some well-defined sense, but much better conditioned. Formally, given A_r as some well-behaved approximation of A, one solves

$$\hat{x}_\lambda = argmin\{\|A_r(x) - b\|_2^2\} \quad (4.9)$$

In the discrete linear case, the most common version of this approach is called the Truncated Singular Value Decomposition (TSVD) and A_r a well-conditioned low-rank least squares approximation to A.

We note that under quite general conditions the linear versions of these two approaches can be shown to be closely related, somewhat surprising given the difference in their conceptual underpinnings. We discuss this in more detail in what follows. In statistical approaches X and n are considered as random vectors with given probability models. In particular, the probability model for X encodes our belief about how a reasonable, or physiological, solution should behave.

Applications of this approach to inverse electrocardiography to date have tended to use a simplified version of this general model with the following assumptions:

1. A linear forward matrix A
2. Gaussian statistical models for both X and n

Under these assumptions, the Bayesian solution becomes the maximum of the posterior probability distribution of $\boldsymbol{x}$, given the measurements $\boldsymbol{b}$, and this solution is commonly called the Maximum a Posteriori (MAP) solution. Which is the best linear mean square error solution for $\boldsymbol{x}$, but not the optimum Bayesian solution.

Independent of the assumption of probability models, there are a number of ways one can treat the temporal variation of $\boldsymbol{x}$ statistically. For simplicity, we discuss the possibilities in the context of Gaussian distributions for both the unknowns and the noise, so we only need to specify the mean and the covariance. The simplest and most widely used model assumes that each time instant is a vector drawn from a stationary distribution. In any of the zero-mean cases, the MAP solution can be written as:

$$\hat{x}_{MAP} = (A^T C_n^{-1} A + C_x^{-1})^{-1} A^T C_n^{-1} b, \quad (4.10)$$

where C_x and C_n are appropriate solution and noise covariance matrices, respectively.

If $\boldsymbol{n}$ is assumed to be white noise ($C_n = \sigma_n^2$), the solution becomes:

$$\hat{x}_{MAP} = (A^T A + \sigma_n^2 C_x^{-1})^{-1} A^T b, \quad (4.11)$$

4.3. Potential-Based Inverse Solutions from Body Surface Measurements

In this section, we describe the conceptually most straightforward version of the inverse ECG problem – reconstructing potentials on the heart surface (denoted HSP in what follows) from measurements on the body surface. the forward transfer between these potentials can be expressed by a matrix. The nature of the matrix involved demands a more elaborate treatment than just an inversion. We begin with the most common variant reconstruction of epicardial potentials – and explain in more detail the two standard deterministic approaches outlined earlier: Tikhonov regularization and TSVD. We then describe some recent efforts to achieve more reliable and accurate solutions by incorporating more prior information through use of truncation of iterative matrix equation solutions, more complicated statistical models, multiple spatial regularization constraints, and constraints on temporal behavior. We then discuss two other potential-based source models to which inverse solutions have recently been applied: transmembrane potentials and then transmural potentials. From here on, we will use $\phi_H(x,t)$ to represent HSPs at positions $\boldsymbol{x}$ on the heart surface at time t, and similarly, $\phi_B(y,t)$ to represent the potential at positions $\boldsymbol{y}$ on the body surface.

By posing the inverse problem in terms of reconstructing heart surface potentials, the problem is linear, and the resulting solution is theoretically unique. However, as already noted, because the inverse problem is ill-posed, such a formulation is inherently unstable; even extremely low levels of signal noise or very small geometric errors can result in an unbounded solution. In order to stabilize the problem and to obtain a reasonable solution, it is necessary to incorporate further constraints before attempting to solve the equations.

The most common and straightforward approach to this problem is via regularization of the inverse problem. A successful regularization procedure will yield a feasible solution that has useful properties in common with the exact solution of the underlying unperturbed problem. The main challenges in applying this approach are to determine both the type and the amount of regularization required to produce the desired solution. Regularization can be viewed as a procedure that imposes constraints on the solution, typically on its magnitude or smoothness. The goal is that these constraints relate to the underlying physiology or other known information about the solution.

Treating every individual column of the systemin (4.5) independently and solving for each column, ϕ_B, to obtain the regularized solution, ϕ of (4.6), i.e., the heart potential $\phi_H(t)$

$$\phi_H = A_{\lambda_t}^{\dagger}\,\phi_B, \tag{4.12}$$

where $A_{\lambda_t}^{\dagger}$ is the regularized inverse matrix at time t.

Next, we describe in detail two of the most common specific approaches to generating the pseudo-inverse, both of which are widely used in cardiac applications.

4.3.1. Tikhonov Regularization

The Tikhonov regularized solution is obtained by minimizing an appropriate objective function

$$\phi_H = min\{\|A\phi_H - \phi_B\|_2^2 + \lambda_t^2\|R\phi_H\|_2^2\}, \quad (4.13)$$

where R is an $N \times N$ constraint matrix and $\|\cdot\|\cdot$ is the Euclidean norm. The R term in (4.13) helps to constrain (or regularize) the inverse solution. The first term in (4.13) represents the least-squares solution to each column of (4.5), while the second term constrains, in the spatial domain, the amplitude of the solution according to the choice of the particular constraint matrix R. The constant λ_t is the *regularization parameter* at each time, t, which controls the weight given to the residual and solution norm and hence controls the degree of smoothing. The full solution is then a balance between the unconstrained least squares solution and a set of constraints that in some way encapsulates a priori knowledge of a physiologically realistic solution.

There are three Tikhonov regularization constraints typically used in inverse electrocardiography, known in that literature as zero-order, first order, and second-order Tikhonov regularization. Zero-order Tikhonov regularization uses $R = I$, the identity matrix, which effectively limits the total magnitude of the solution; first-order Tikhonov regularization uses $R = G$, a discrete approximation to the surface gradient operator, and limits the steepness of the solution; and the second-order Tikhonov method uses $R = L$, a discrete approximation to the surface Laplacian operator, to restrict the rate of change of the steepness, i.e., the overall nonsmoothness of the solution. With the Tikhonov approach we can write a closed form solution for (4.13) that leads to an expression for the pseudoinverse.

$$A_{\lambda_t}^{\dagger} = (A^T A + \lambda_t^2 R^T R)^{-1} A^T, \quad (4.14)$$

Substituting (4.14) into (4.12), the solution for zero-order Tikhonov regularization becomes

$$\phi_H = (A^T A + \lambda_t^2 I)^{-1} A^T \phi_B, \quad (4.15)$$

4.3.2. Truncated SVD (TSVD)

Another method of treating the ill-conditioned nature of the transfer matrix A is to derive a new problem with a well-conditioned *rank deficient* transfer matrix. A common approach is to use the singular value decomposition (SVD). From a singular value decomposition, one can determine the *principal components* of the information contained within a matrix.

If an SVD is applied to a transfer matrix A, then the transfer matrix can be written in the form

$$A = U_A \sum_A V_A^T = \sum_{i=1}^{N} \sigma_i u_i v_i^T \quad (4.16)$$

Where U_A and V_A are orthogonal matrices (their columns are orthonormal) and diagonal matrix $\sum_A$ contains as its diagonal entries the *singular values*, while u_i and v_i are the column vectors forming these orthogonal matrices and σ_i are the singular values for

$i = 1, \ldots, N$. We assume in this case that the number of measurements, or the row-size of A, M is larger than the number of HSP reconstruction sites, or the column size of A, N.

If $M < N$ some details change but the results are the same. The singular values are greater than or equal to zero and are typically sorted in order of decreasing size. For a matrix representing an ill-posed process (as is the case here) the range of nonzero singular values covers many orders of magnitude.

The smaller singular values, the reciprocals of which are used in the least-squares inverse, will tend to magnify components in the data. While data components in the subspace of the singular vectors corresponding to the larger singular values are not amplified that dramatically. Thus, these "low singular value" components dominate the solution. Moreover, for forward solutions that smooth high frequencies, as in electrocardiography, those same singular vectors will also represent high frequencies and thus contain only highly attenuated information about the sources. The result is extreme noise amplification, as seen in our original example in Fig. 4.1.

To overcome this problem, the truncated SVD(TSVD) solution proposes to substitute for A a low rank approximation, which simply leaves out the modes with unstably small singular values, i.e., we replace A with

$$A_{\lambda_t} = \sum_{n=1}^{\lambda_t} \sigma_{A(n)}\, u_{A(n)}^T\, v_{A(n)}, \quad \lambda_t \leq N, \tag{4.17}$$

where $u_{A(n)}$ and $v_{A(n)}$ are the nth vectors from the SVD of A, $\sigma_{A(n)}$ are their corresponding singular values, and N is the full rank of the matrix. The size of λ_t determines the level of regularization for time t (where λ_t is a positive integer). We wish to emphasize that the right singular vectors have the dimension of the data space, while the left singular vectors have the dimension of the solution space.

The TSVD solution is obtained by minimizing the objective function

$$\phi_H = min\left\|A_{\lambda_t}\phi_H - \phi_B\right\|, \tag{4.18}$$

The solution to this can be computed by means of the pseudo-inverse

$$A_{\lambda_t}^{\dagger} = (A_{\lambda_t}^T\, A_{\lambda_t})^{-1}\, A_{\lambda_t}^T, \tag{4.19}$$

The pseudo-inverse can be found easily from the SVD of A as

$$A_{\lambda_t}^{\dagger} = \tilde{V}_A \tilde{\sum}_A^{-1} \tilde{U}_A^T, \tag{4.20}$$

where $\tilde{U}_A$ and $\tilde{V}_A$ contain the first λ_t columns of U_A and V_A respectively,

and $\tilde{\sum}_A$ is a $\lambda_t \times \lambda_t$ diagonal submatrix of $\sum_A$.

Substituting (4.20) into (4.19), the TSVD solution is

$$\begin{aligned}\phi_H &= \widetilde{V}_A \widetilde{\Sigma}_A^{-1} \widetilde{U}_A^T \phi_B \\ &= \sum_{n=1}^{\lambda_t} \frac{u_{A(n)}^T \phi_B}{\sigma_{A(n)}} v_{A(n)}(x)\end{aligned} \quad , \tag{4.21}$$

It is useful to compare this TSVD solution with the zero-order Tikhonov solution given in (4.15). Expressing (4.15) in terms of the u and v components from the SVD of A we can write

$$\begin{aligned}\phi_H &= \left(A^T A + \lambda_t^2 I\right)^{-1} A^T \phi_B \\ &= \sum_{n=1}^{N} f_n(t) \frac{u_{A(n)}^T \phi_B}{\sigma_{A(n)}} v_{A(n)}(t)\end{aligned} \quad , \tag{4.22}$$

where $f_n(t)$ are the Tikhonov filter factors given by

$$f_n(t) = \frac{\sigma_{A(n)}^2}{\sigma_{A(n)}^2 + \lambda_t^2} = \begin{cases} 1 & \sigma_{A(n)}^2 \gg \lambda_t \\ \frac{\sigma_{A(n)}^2}{\lambda_t^2} & \sigma_{A(n)}^2 \ll \lambda_t \end{cases} , \tag{4.23}$$

These filter factors have the effect of filtering out contributions to the reconstructed ϕ_H that correspond to the small singular values while leaving the SVD components corresponding to large singular values almost unaffected. When the filter factor equals zero or one, Equations (4.21) and (4.22) are identical.

4.3.3. Truncated Iterative Approaches

Truncated iterative methods are another deterministic approach to solving the inverse problem. In these methods, a sequence of candidate solutions is produced and each one is evaluated according to a "goodness" criterion: if the solution meets some threshold of accuracy, the iterations stop and otherwise they continue with additional candidate solutions. These approaches draw on the standard techniques for solving large linear systems of equations. However, because such methods will converge to unreliable solutions, the regularization approach is to truncate the iterations before convergence. In this case, the number of iterations plays the role of a regularization parameter.

Iterative methods are especially favorable for large-scale problems (where direct regularization methods are computationally expensive). And in problems in which a matrix representation of the forward operator or an explicit representation of the inverse solution are not available However, their use is not limited to these cases; they can be used as an alternative approach to direct regularization approaches. Most iterative methods can be represented in terms of filter factors such as those already discussed for Tikhonov and TSVD approaches. which demonstrates the similarity between iterative methods and direct methods in their filtering of the small singular values of the forward matrix. In regularized iterative methods, the

solution converges to the lower frequency modes of the right singular vectors of A in the earlier iterations, and thus stopping the iteration filters out the effect of the otherwise amplified higher frequency modes.
The most commonly used iterative methods in this context are in the class of Krylov subspace methods. An early attempt of this approach was reported in which the Conjugate Gradient (CG) iterative method was combined with regularization constraints.
More recently, scientists have reported using another Krylov subspace method, the Generalized Minimum Residual (GMRes) method, with significant success. Reports suggest that the inverse solution reconstructed by the GMRes method was more accurate than the Tikhonov solution in terms of the pattern and localization of epicardial potentials.

However, despite the strong empirical evidence supporting the GMRes approach, there is no theoretical justification to explain why this method reconstructs a localized solution better than Tikhonov regularization. We note that in both approaches, Tikhonov-type regularization was combined with the iterative regularization, either by iterating on a regularized set of equations or by using the Tikhonov solution as the starting point for the iterative regularization.

4.3.4. Statistical Approaches

Although some of the original reports on inverse electrocardiography for epicardial potentials were based on the statistical model in (4.10), the main challenge of this approach remains specifying the model. The simplest technique, which was adopted by these early investigators, is to assume that not only is the noise white and uncorrelated, but that the solution has the same structure, with zero mean and variance σ_x^2. In this case the relevant equation becomes

$$\hat{x}_{MAP} = (A^T A + \frac{\sigma_n^2}{\sigma_x^2} I)^{-1} A^T b \qquad (4.24)$$

This is the same as Tikhonov regularization with an identity matrix as the regularizer and with a statistical model for picking the regularization parameter. Since a good model for the variance of the unknown solution is hard to specify, most of the research that followed these early reports concentrated on the deterministic regularization approaches described above.

4.3.5. Multiple Spatial Regularization Operators

All of the traditional methods have produced results that showed that it is indeed possible to recover meaningful information about cardiac surface potentials from body surface measurements. However, these results were neither precise nor reliable enough to be really attractive for potential clinical or even scientific use. One approach that has shown some success is to combine two distinct types of regularization.

Another approach starts from the observation that if the problem is that we need constraints to improve inverse solutions because the measurements contain insufficient information in themselves, one can consider the possibility of using more than one constraint at the same time. This approach was applied to inverse electrocardiography using both Tikhonov zero-order and Tikhonov second-order regularizers. The results were reported to be somewhat improved, and in particular were more robust to the exact choice of regularization parameters. This approach had its difficulties, however, in particular in picking a good set of regularization parameters without undue computational burden, and this problem gets dramatically worse if the number of constraints increases beyond two. At the same time, it was not clear that any particular pair of constraints contained the most useful information for reconstructions.

In this approach, the optimization algorithm is stopped as soon as one finds a solution that matches all the constraints similar in spirit to the way in which the iterative Krylov-subspace linear system solvers like CG or GMRes are truncated when dealing with ill-posed problems. The application of this approach to inverse electrocardiography, using one such convex optimization algorithm called the ellipsoid algorithm. However, again there was no clear set of constraints on the potentials that emerged as giving the level of accuracy and reliability required.

4.3.6. Spatio-Temporal Approaches

The temporal behavior of cardiac electrical signals is known to have a strong deterministic component. It would make sense to make use of this strong prior knowledge to constrain solutions over multiple time instants. One approach involves parameterizing the sources by means of fiducial time parameters. With potential-based methods, the temporal behavior of the source would also seem to be a powerful tool in the inverse solution toolbox.

In fact, from a statistical standpoint, it is commonplace in the fields of statistical signal processing and estimation theory that, if the quantities in (4.10) are taken as potentials at single time instants, the solution is optimal only if the desired heart surface potentials are temporally uncorrelated. Any reasonable model for cardiac sources will be far from temporally uncorrelated, and thus it is not surprising that inverse electrocardiography algorithms have been introduced to take advantage of models of expected temporal behavior.

We first treat extensions of the *heart surface potential* (HSP) inverse solution formulation described earlier to include temporal or spatio-temporal constraints. We then discuss a relatively new approach that replaces the heart surface model with a *transmembrane potential* (TMP) model. Based on this model, temporal constraints that are a reasonable approximation of TMP temporal behavior can be applied; the hoped-for advantage is that TMP temporal behavior may lend itself to simpler, more effective and useful, constraints.

4.3.6.1. Spatio-Temporal Constraints with Heart Surface Potential Models

Some of the earliest work in this field used a simple on–off model to incorporate temporal information, yet this wasn't pursued due to various factors (such as regularization constraints and parameters and the development of appropriate validation models). Later on, researchers began to consider again the use of temporal behavior. And by simply adding a temporal constraint after an initial Tikhonov solution, this method penalized subsequent time instants for changing too dramatically from the preceding one.

Formally, (4.7), with a linear model and linear regularization constraint, becomes:

$$\hat{x}_\lambda = argmin\{\|Ax - b\|_2^2 + \lambda^2 \|R(x - x_0)\|\} \tag{4.24}$$

where x_0 represents an initial estimate of the solution. Thus, this approach requires an initial estimate of the solution; the quality of the resulting inverse solution was shown to depend on the quality of this estimate, and the final solution is therefore biased towards it.

Starting from this approach, several alternative methods were introduced that attempted to combine spatial and temporal constraints into a true spatio-temporal regularization method. The three methods that had the most success was:

1. A method that added a temporal constraint to one (or more) spatial constraint(s).
2. A method that added an equation containing an explicit model of temporal evolution of the potentials to the standard forward problem equation (which models the spatial relation between potentials on two different surfaces), with the pair of equations formulated as a state-space model and solved via a Kalman filter or smoother.
3. A method which attempted to simultaneously regularize in time and space by finding a temporal whitening transform according to a statistically informed assumption. The initial version of the method was presented in a specific variant with a heuristic justification that Greensite has since developed into a more general and theoretically based formulation using an assumption he called "isotropy". Indeed, this earlier method has become perhaps the most widely accepted for reconstruction of potentials by inverse electrocardiography.

Initially, it appeared that these were three distinct methods with no clear connection or comparison among them except in terms of exemplary results. However, recently it has been shown that in fact all the three of these methods can be cast into the same statistical framework. We outline this framework here, leaving the details to the literature, and then proceed to describe in some detail the popular implementation of Greensite's method. The key to unifying these approaches is to consider (4.11) with the quantities defined as in (4.3) and (4.4); the vectors are all block vectors concatenating the spatial distribution of the potentials across all time instants, and $\mathcal{A}$ is a block diagonal matrix with the forward solution matrix repeated on the diagonal blocks. We employ the statistical model described earlier in which both the noise and the unknown signals are assumed to be zero-mean, and the noise is assumed to be white as noted, here in

both space and time. In this case the key quantity that governs the solution is the spatio-temporal covariance matrix C_x, or equivalently, its inverse:

$$C_x = E\{x\, x^T\} = E\left\{\begin{bmatrix} x(t_0) \\ x(t_0+1) \\ \vdots \\ x(t_n) \end{bmatrix} [x(t_0)^T \quad x(t_0+1)^T \quad \dots \quad x(t_n)^T]\right\}, \tag{4.24}$$

where $E\{\cdot\}$ represents statistical expectation. Thus, in the spatio-temporal formulation, C_x is a spatio-temporal covariance matrix which can be divided into blocks of the form $E\{x(t_i)\, x(t_j)^T\}$. The dimension of each block is the number of nodes used on the heart surface; the number of blocks (in both directions) is equal to the number of time instants; and each block is a spatial cross-covariance matrix at different time instants. The temporal covariance is embedded in the variation from block to block across the matrix, as indexed by the time lag between t_i and t_j.

Considering this formulation, two significant problems arise:

1. How to derive or estimate the parameters required to populate this matrix?
2. How can one handle the dramatic increase in computational requirements stemming from the fact that the matrices now have dimensions on each side that are equal to the number of spatial nodes on the relevant surface multiplied by the number of time instants considered?

Each of the three methods introduced separately – multiple regularization, Kalman filtering, and isotropy – solves exactly this same pair of problems, but in different ways, based on different assumptions. Each assumption turns out to impose a particular structure on C_x or its inverse, which then leads to a specific algorithm or set of algorithms.

Here we briefly describe the simplest form of the approach taken in each method.

1. *Multiple regularization*. In this approach we assume that we can build separate constraints for the spatial behavior and for the temporal behavior. When formulated in terms of statistical regularization, such assumptions impose a particular structure on the inverse of the spatio-temporal covariance matrix, as the sum of two block matrices; one is a block diagonal matrix, which contains the spatial regularization constraints, and the other is a block matrix that has all its blocks restricted to be themselves diagonal matrices, and that enforces the temporal regularization constraint. A consequence of this structure is that fast algorithms can be developed which reduce the computational complexity down to the order of the number of nodes in the heart surface model.
2. *State-space model.* In this approach, the spatial distribution of potentials at each time instant is modeled by a prediction or temporal evolution equation, in which the potentials from each time instant evolve from the previous time by applying a known prediction matrix plus a random perturbation. For example, imposing an identity prediction matrix assumes that the potentials do not change rapidly from one time sample to the next (a random walk model). More general models, which consider, for instance, the behavior of neighboring nodes, can easily be employed. One consequence of these assumptions is that the resulting structure of the spatio-temporal covariance matrix shows a kind of block exponential decay (for a stable model) across the blocks. More

importantly, the *inverse* covariance matrix has a block tri-diagonal structure; in other words, regularization is applied only between adjacent time instants in this model. The computational consequence of this simple structure is in fact the Kalman filter/smoother algorithm, and thus, a reduction in computational complexity is achieved.

3. *Isotropy model*. One consequence of the assumption of isotropy is what in statistical terms is known as separability between the spatial and temporal correlation. The effect on the structure of the inverse covariance matrix is that all are exactly the same except for a scalar multiplication. Mathematically, this assumption means that the spatio-temporal covariance matrix can be factored as the Kronecker product of a single spatial covariance matrix by a single temporal covariance matrix. As a further consequence of this assumption, the temporal covariance of the *heart* surface potentials is identical to within a scalar to that of the *body* surface potentials. This assumption suggests the following algorithm:
    (a) Estimate the temporal covariance matrix of the body surface potentials directly from the data.
    (b) Find the SVD of this matrix; its right singular vectors can be used to decorrelate, or whiten, the entire problem, by a simple matrix multiplication.
    (c) Once the problem has been so whitened, it is optimal to solve it column by column. Spatial regularization is still required and can be done by any relevant method, including Tikhonov regularization or TSVD.
    (d) Once the column-by-column solutions are complete, re-correlate by multiplying by the transpose of the decorrelating matrix. Since this matrix comes from the SVD and is therefore orthogonal, the optimality of the solution is preserved.

    This algorithm reduces the computational complexity to the order of the spatial dimension.

Each of these methods has its advantages and limitations. The multiple regularization
approach gives the designer great freedom in choosing regularizers but requires considerable prior knowledge to make good choices. Moreover, we cannot consider nearby spatial nodes when regularizing in time without incurring a dramatic increase in computational complexity.

The state-space model gives even more flexibility in the design of the temporal model, and indeed, this model has a very direct physical or physiological interpretation as the expected temporal progression of the solution. Moreover, it opens access to a truly vast literature of modeling and solution methods. However, it does depend on an effective prior knowledge of this evolution, and most applications of this method to inverse electrocardiography have simply used the random walk approach. In addition, it does not allow one to directly consider more than one previous time instant in predicting the next time instant, again unless one is willing to increase the computational order and the modeling complexity.

The isotropy/separability method has the significant attraction that one does not need to model or make any further explicit assumptions on the temporal behavior However, the underlying assumption of isotropy, or separability, is difficult to interpret in physical or physiological terms, and it is not yet clear how

one can verify or test its accuracy. Moreover, the assumption rests on the idea that one knows nothing about the expected temporal behavior of the solution. If, in fact, one has some prior knowledge, it is not clear if or how to incorporate that knowledge into this approach. Nonetheless, as already noted, not only the original authors of this method but also independent groups have reported success with it in several settings.

There is one key aspect of the isotropy method, which is critical to its performance After the decorrelating transform achieved by multiplication of the data by the right singular vectors of the data matrix, one has exchanged the original set of L equations, where L is the number of time instants used, for a new set of L equations. each of these equations still involves the badly conditioned forward matrix A and thus still needs to be spatially regularized. If one uses a standard method to find an appropriate regularization parameter for each equation and then solves all L equations, the results are dominated by noise and are not useful. This is true because the new set of equations has its own set of "singular values" in the transformed domain, which are unknown even under the isotropy assumption Therefore, one must find a way to truncate this process, only solving the set of equations that contain reliable information included an ad hoc method for exactly this truncation, via a kind of "double TSVD," which uses an SVD of the data matrix to determine which equations to solve, and an SVD of A to regularize each of those solutions.

Proceeding to a detailed description of the version of the isotropy method generally used in practice. We start by performing an SVD of the torso surface signal data matrix ϕ_B, where ϕ_B is a specific variant of the matrix B in (4.5) which stores the signal information in its rows and the potentials at each time instant in its columns, to determine its principal components. The result of this process is a factorization of ϕ_B into a matrix of spatial singular vectors U_B and temporal singular vectors V_B. Because of the spatial smoothing and attenuation, only a relatively small number of time signals can represent almost all the information in this data matrix; the rest of these components are dominated by noise.

We can basically filter the data matrix to remove the noise-dominated components before attempting an inversion this might improve the results. We note that this relatively low-dimensionality of the torso data has been known, and used, for a long time.

Specifically, the columns of UB and VB corresponding to small singular values are assumed to correspond to noise dominated signals and are typically removed. A determination of which singular values correspond to this noise subspace can be estimated in a somewhat ad hoc fashion by plotting the log of the ordered singular values against the rank of the matrix obtained from the appropriate subsets of these singular values. This typically results in a characteristic 'L' shaped curve as shown in Fig. 4.4. It has been postulated that the singular values below the point at which the curve levels off correspond to the noise subspace. A number of different methods have been used for determining the best place to locate this leveling off point. However, none of these has been found to be truly reliable in practice.

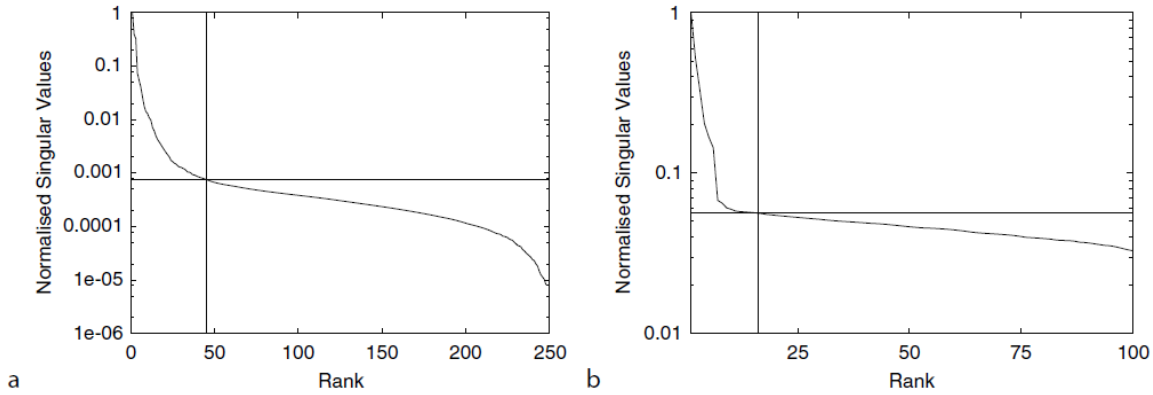

Figure 4-3 Singular value spectra of two different torso surface signal matrices

Singular value spectra of two different torso surface signal matrices. In panel (a) are the singular values from a signal matrix recorded from a normal male volunteer and in (b) signals from a simulated cardiac source. The singular values have been normalized and plotted on a log scale to emphasize the curvature in their values. The lower right quadrant.

The key idea of this version of the isotropy method is not to simply filter the data, as shown, but rather to combine this filtering with regularization. To start, we write the SVD expansion of the matrix of torso potentials matrix as

$$\phi_B = U_B \sum_B V_B^T, \tag{4.27}$$

where U_B is an $M \times Ne$ matrix that represents the spatial basis, $\sum_B$ is an $Ne \times Ne$ matrix with the singular values stored on the diagonal, V_B is an $Ne \times Ne$ matrix that represents the temporal basis, and $Ne = min(M, L)$ where M is the number of recording electrodes on the body surface and L is the number of time samples recorded. We assume once more that the number of measurement locations is greater than the number of nodes on the HSP model.

Inserting (4.27) into (4.12) leads to

$$\phi_H = A_{\lambda_t}^{\dagger} \phi_B = A_{\lambda_t}^{\dagger} U_B \sum_B V_B^T , \tag{4.28}$$

Multiplying on the right by V_B, which is what the isotropy assumption would lead to as the optimal decorrelating transform, leads to

$$\phi_H V_B = A_{\lambda_t}^{\dagger} \phi_B V_B = A_{\lambda_t}^{\dagger} U_B \sum_B V_B^T V_B = A_{\lambda_t}^{\dagger} U_B \sum_B \tag{4.29}$$

Next, we define

$$\Gamma = \phi_H V_B \tag{4.30}$$

By Inserting (4.29) into (4.30), we have

$$\Gamma = A_{\lambda_t}^{\dagger} \phi_B V_B = A_{\lambda_t}^{\dagger} U_B \sum_B \tag{4.31}$$

This problem is tackled for each column vector γ_i of Γ, paired to each column vector i of matrix $U_B \sum_B$, with $i = 1, \ldots, N_e$.

We use the noise subspace method, which permits the selection of regularization parameters for each value of i in any type of pseudo-inverse selected. The solution to (4.27) is then found from

$$\phi_H = \Gamma V_B^T \tag{4.32}$$

4.4.6.2 Spatio-Temporal Constraints with Surface Transmembrane Potential Models

transmembrane potentials can be used as the source strength of a double layer for an appropriate forward model to be formulated. In such models, the forward solution maps the TMPs directly to the body surface potentials. TMPs have some disadvantages in comparison with HSPs; for example, they require additional isotropy and homogeneity assumptions, and are only guaranteed to be unique if these assumptions are valid. TMPs are more difficult to measure directly in experimental settings, especially for purposes of validation of inverse solutions.

4.4. Time Based Inverse Algorithms

In HSP and TMP methods, the potentials at each point in space and time are essentially free variables thus, strong explicit constraints are required to achieve useful solutions. Note that it' s not a simple task to find constraints that both capture relevant physiology and are mathematically tractable.

Another general approach to the inverse electrocardiography problem imposes very strong physiologically based constraints in the source model itself and reduces the number of free parameters. The basic idea is to focus on reconstructing the time of activation at each point in space on the epicardial and endocardial surfaces. A benefit of this approach is that from a physiological standpoint, activation times are the most clinically relevant aspect of cardiac electrical activity. it is simpler to reconstruct them directly. this formulation becomes much better constrained; instead of needing to reconstruct an entire time series at each surface location, we only need to reconstruct one number at each location to capture the entire heartbeat. One particular advantage for this approach is greater robustness to errors in the geometry of the forward model.

The main drawback of the activation time imaging formulation is that the forward problem, which now maps activation times to body surface potentials, is nonlinear. Hence, nonlinear optimization algorithms

must be used to solve the inverse problem for the set of activation times that best fits the body surface data.

A powerful algorithm by Huiskamp and Greensite based on activation imaging approach has re-invigorated this area of investigation. This algorithm is centered around the identification of what are known as "critical points" and associated times in the surface activation function which are found using a modified Multiple Signal Classification algorithm.

The key idea behind this critical-point approach stems from the observation that when an evolving cardiac excitation wavefront intersects the epicardial surface, a hole develops in the wavefront. This is a significant change to the topology of the wavefront, which generates a sudden alteration in the surface potential recordings. If $\tau(x)$ is defined to be the activation time on the surface of the heart, where x is a location parameter on the heart surface, and $\tau(.)$ has units of time.), these breakthrough points are critical points of $\tau(x)$, that is, $\nabla \tau(x') = 0$, where x' is the location of the breakthrough point.

This critical point observation leads to the two following important results:

1. x' is a critical point of $\tau(x)$ (with critical time $\tau(x') \Leftrightarrow a$) is in the space spanned by the spatial singular functions of ϕ_B, where a is the column of the V_m to ϕ_B transfer matrix A corresponding to x'.
2. with all critical points of $\tau(x)$ determined, the computation of $\tau(x)$ is a well-posed problem on both the epicardial and endocardial surfaces.

The key assumption required to prove the first point is that $\mathrm{V_m}$ is modeled as a uniform step jump across the wavefront.

$$V_m = a + bH\big(t - \tau(x)\big), \tag{4.33}$$

where a is a constant offset value, b controls the height of the step jump of the action potential, H is the Heaviside step function, and $t \in [0, T]$. The Heaviside step function models an elementary source at a point on the heart surface x, excited at time $\tau(x)$.

The previous equation only represents phase 0 of the action potential and hence the inverse algorithm is specific to the intervals occupied by the **P** wave (atrial depolarization) and the **QRS** complex (ventricular depolarization) of the ECG. The equation be;

$$V_m = a + b - bH\big(t - \tau(x)\big), \tag{4.34}$$

Consider $\tau(x)$ to be a time corresponding to the phase 3 (repolarization) of the action potential, then results similar to the two listed earlier could be obtained. Thus, any inverse algorithm derived from those two results for imaging depolarization could be used for image repolarization, and thus this approach is a means of complete fiducial-time based imaging. repolarization occurs over a relatively long time period, and therefore, it is

difficult to define a distinct recovery time (compared with the activation time), which means that (4.34) is a more severe approximation than (4.33). Moreover, since it is the change in time of the TMP that produces the "source" in this model, the effective signal-to-noise ratio is considerably reduced. Also, repolarization occurs during the mechanical motion of the heart, meaning that the transfer matrix may contain substantially more geometric error in the repolarization phase compared with the depolarization phase.

The use of the critical-point method to locate breakthroughs suggests a two-step approach for fiducial-time based imaging:

1. Find the critical points and times using the theory above, then
2. Obtain the entire activation sequence by solving a non-linear minimization problem in which the objective is to minimize the difference between the calculated torso potentials $\hat{\phi}_B$, and the measured potentials, ϕ_B (where $\hat{\phi}_B$ is calculated using $A(x, y)$ and any estimate of $\tau(x)$).

During this optimization process, the critical points and times that have been identified can either be fixed or constrained to remain local extrema of $\tau(x)$. Additional constraints on the optimization process can be imposed, such as the surface Laplacian of $\tau(x)$.

$$E(\tau(x)) = \min\{\|\phi_B - \hat{\phi}_B\|_2 + \lambda . L\tau(x)\}, \tag{4.35}$$

where E is the objective function, λ is a parameter controlling the degree of regularization, and L is a discrete approximation of the Laplacian of the excitation field.

The following algorithm computes the critical points and times. First, the signal matrix, ΦH, recorded from surface electrodes, and $A(x, y)$ mapping from V_m to ϕ_B, are required. The spatial singular functions of $\phi_B(y, t)$ are computed using singular value decomposition.

$$\phi_B = U_B \Sigma_B V_B^T \quad \text{with an effective R rank} \tag{4.36}$$

Noting that each column of A corresponds to the map from a particular node x on the heart surface to all the body surface nodes, the following function can be defined for each x for all values of $t \in [t_0, t_1]$

$$M_{t_0}^{t_1}(x) = (1 - \sum_{r=1}^{R}[\tilde{a}(y) \cdot u_r(x)^2])^{-1} \quad (4.37)$$

where $\tilde{a}(y) = \frac{a(y)}{\|a(y)\|}$ is the unit Euclidean normalized column of the transfer matrix $A(x, y)$, $u_r(x)$ is the rth column of the spatial left singular matrix U, and R is the number of singular values retained. The summation term is the projection of the normalized vector $\tilde{a}(y)$ onto the vector space spanned by the spatial singular vectors $u_r(x)$. If $\tilde{a}(y)$ is contained in this space then the summation will yield 0 (and hence $M_{t_0}^{t_1}(x)$ will be infinite), while if no component of $\tilde{a}(y)$ is contained in this space, the summation will yield 1. This is the principle behind the well-known MUSIC method for spectral estimation and array processing.

Once the critical points or initial activation times are found, we are left with a nonlinear optimization problem. With the critical points serving either simply as a starting point for the optimization or also as an additional constraint throughout the procedure. We note again that this problem is nonlinear because $\hat{\phi}_B$ depends on the fiducial times in a highly nonlinear manner, in contrast to its dependence on the HSP or TMP values

$$Z_{xt} = M_0^t(x) - M_t^T(x) \quad (4.38)$$

In the following figure, we show sample functions defined by the rows of Z_{xt} (called the critical point functions). In practice, the distinction between critical points on the one hand, and activation times on the other, is not clear, due to noise and errors. Once the critical points or initial activation times are found, we are left with a nonlinear optimization problem. Note that this problem is nonlinear because $\hat{\phi}_B$ depends on the fiducial times in a highly nonlinear manner, in contrast to its dependence on the HSP or TMP values.

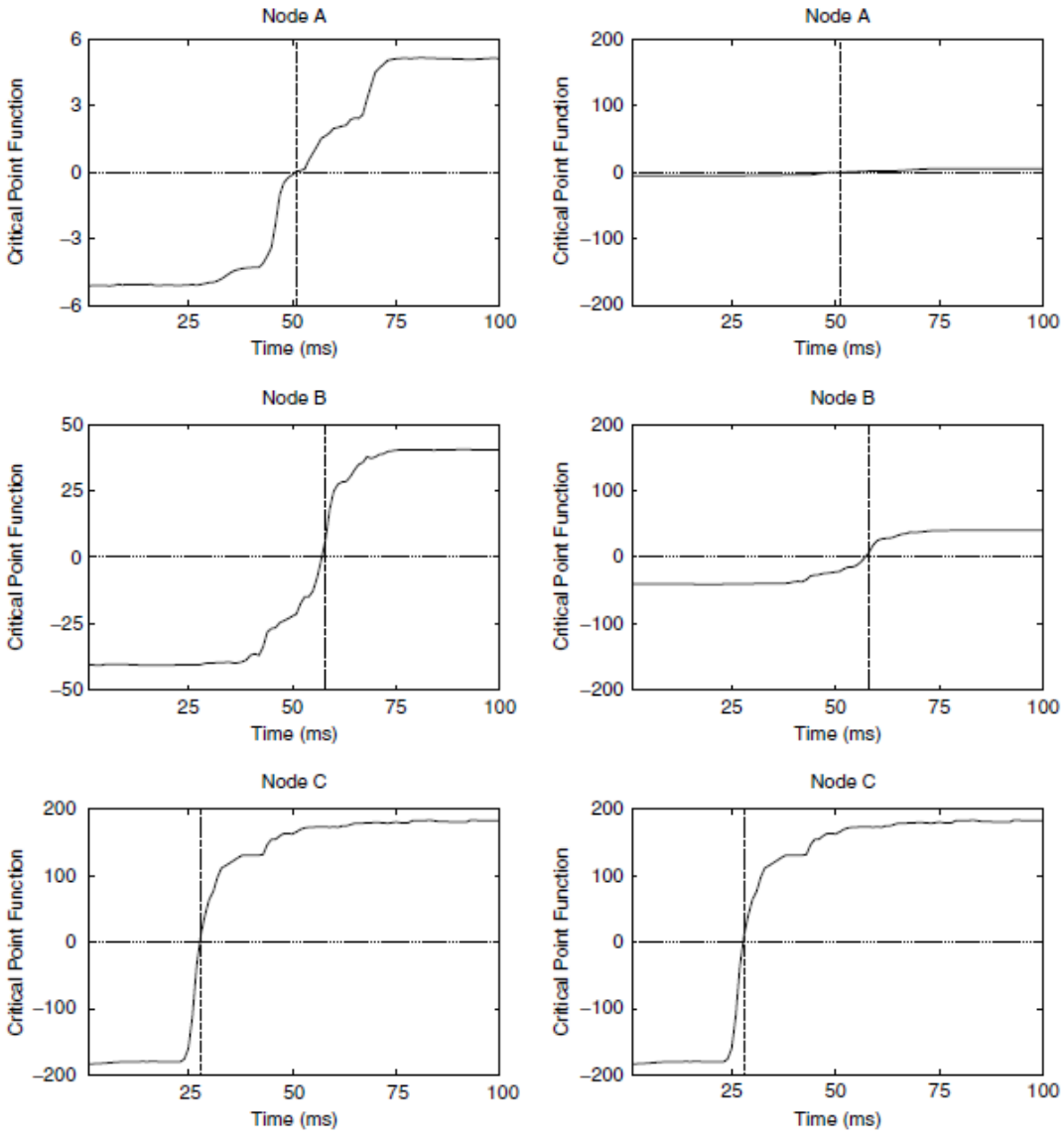

Figure 4-4 Comparison of critical point functions between critical and noncritical points

Figure 4.4 Comparison of critical point functions between critical and noncritical points. The left *column* shows the three functions and their corresponding ranges. The right column re-plots all the functions scaled to the same range. The larger jump in the critical point function for node C indicates that it corresponds to a critical point. The vertical line marks the points at which the functions cross the *horizontal axis* (critical times) and provides an estimate of the activation time at each location.

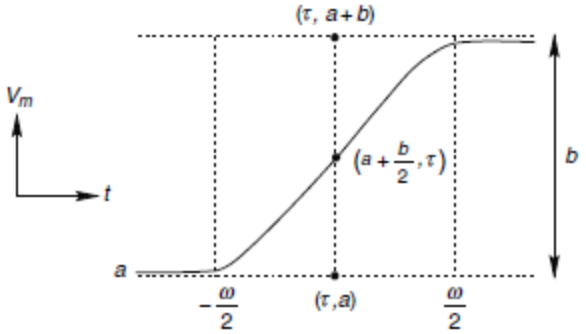

Figure 4.5 Approximation of the Heaviside step function by a sigmoidal function. This function contains the parameters ω and b to determine its shape. The ω parameter controls the width or window of the activation upstroke (i.e., the duration of the depolarization), while the b parameter controls the height of the transmembrane jump, i.e., the difference between the resting potential (a) and the peak of the action potential upstroke. Time is shown on the horizontal axis and the transmembrane potential on the vertical axis.

All of these functions rely on assumptions about the shape of the action potential that are likely not a practical restriction for normal hearts, but do provide a measure of the expected spatial resolution of this activation-inverse approach.

4.5. Inverse Solutions from Intracavitary Measurements

Another method of inverse electrocardiography is to measure the potentials within a chamber of the heart by means of a noncontact, multielectrode catheter array and then to estimate the endocardial surface potentials from them. The forward model of this approach has a similar form to other potential-based approaches and relates the potentials on the endocardial surface to the measurement leads inside the cavity. Since the measurements are recorded relatively close to the endocardial surface, the corresponding forward matrix is generally less ill-conditioned than the epicardial potential forward matrix. However, obtaining accurate geometrical data from which to construct the transfer matrix is challenging, because it requires medical imaging and because the probe moves due to blood flow in the chamber. Overcoming this challenge is critical, because the accuracy of the computed inverse solutions is highly sensitive to errors in the geometry. The only currently available commercial device based on inverse electrocardiography, the Ensite system, obtains the geometry of the endocardial surface point by point using a catheter navigating on the endocardial surface. The sampled points are then fitted to a bicubic spline surface to improve the estimate of the actual chamber.

Due to the similarity of the forward model of this approach to the potential-based approach, all inverse methods mentioned in Sect. 4.3 apply to this problem as well. Here we briefly describe two approaches that are slightly different from those covered in Sect. 4.4. While initially developed for the endocardial problem, there is no reason why these approaches could not be used for other potential-based formulations.

1. The first approach uses the first order Tikhonov regularization method and the BEM method to define the regularization matrix as the derivative of the endocardial potentials in the normal direction to the surface rather than tangential to the surface. This means that the regularization term constrains the magnitude of the normal current density on the endocardial surface.

2. The second approach is again in the framework of Tikhonov regularization but tries to define a regularization matrix that will satisfy the discrete Picard condition. The regularization matrix in this approach has the same right and left singular vectors as the forward matrix, but its singular values are determined from the energy spectrum of the data. Therefore, the singular values of the regularization matrix and the particular measured data directly affect the filter factors of the inverse solution. The inverse solution obtained by this method produces superior results than the zero order Tikhonov solution.

5. Chapter 5: Overview on Artificial Neural Networks

5.1. What Are Artificial Neural Networks?

5.1.1. Definition

Artificial neural networks are computing systems loosely modeled after the Neural Networks of the human brain. Though not as efficient, they perform in similar ways. The brain learns from what it experiences, and so do these systems. Artificial neural networks learn tasks by comparing samples, without specifically assigned goals.

For example, while learning image recognition, neural networks in training would learn to identify images containing dogs by examining sample images that have been tagged with "dog" or "no dog" labels and then use those results to locate and identify dogs in new images. These neural networks start from zero, with no data about dog characteristics, such as tails, ears, and fur. The systems develop their own understanding of relevant characteristics based on the learning material being processed. (The human brain does not start from zero. Room for a little evolution?)

One significant advantage of neural networks is their ability to learn in nonlinear ways. This means they have the ability to spot features in an image that are not obvious. For example, when identifying oranges, neural networks could spot some in direct sunlight and others in the shade on a tree, or they might spot a bowl of oranges on a shelf in a picture with a different subject. This ability is the result of an activation layer designed to highlight the useful details in the identification process.

An artificial neural network uses a collection of connected nodes called artificial neurons – a simplistic imitation of biological neurons. The connections are versions of synapses and operate when an artificial neuron transmits a signal from one to another. The artificial neuron that receives the signal can process it and then signal artificial neurons connected to it.

There are six types of neural networks, but two are the most popular: Recurrent and feedforward. A feedforward neural network sends data in one direction only. Data moves from input nodes, through hidden nodes (if any exist), and to the output nodes. Feedforward neural networks do not use loops or cycles and are considered the simplest type of neural network. This type of system can include many hidden layers.

A recurrent neural network, on the other hand, uses connections between nodes to create a directed graph as a sequence, allowing for data to flow back and forth. The network forms a directed cycle, which is expressed as "dynamic temporal behavior" – data whose state changes over time and obeys differential equations using time derivatives. The recurrent neural network can use its internal memory to process the sequence of inputs. This type of neural network is popular for handwriting and speech recognition

5.1.2. Architecture of a Neural Network

There are five recognized types of neural networks.

- Single-layer feed-forward network.
- Multilayer feed-forward network.
- Single node with its own feedback.
- Single-layer recurrent network.
- Multilayer recurrent network.

5.1.3. Types of Artificial Neural Networks

Neural networks are synonymously considered as deep learning, inclusive of many layers between input and output, also called hidden layers. An AI engineer masters all the networks in order to succeed in tricky computation situations. Variations on the classic neural network design allow various forms of forwarding and backward propagation of information among tiers.

Types of ANN are:

- FEED-FORWARD NEURAL NETWORKS
  It is the simplest variant of neural networks as it passes information in one direction through input nodes until it makes it to the output node. These are used in technologies such as facial recognition and computer vision.

- RECURRENT NEURAL NETWORKS
  In this model, each node acts as a memory cell & remembers all processed information in order to reuse it in the future. It is used in text-to-speech conversions where it self-learns and continues working towards correct prediction during backpropagation.

- CONVOLUTIONAL NEURAL NETWORKS
  Text digitization, signal processing, NLP, and image classification are some of the advanced applications of CNN as it uses a variation of multilayer perceptron & creates feature maps that record a region of the image which is broken into rectangles.

- DECONVOLUTIONAL NEURAL NETWORKS
  As the name suggests, it is a revered CNN model and aims to find lost features or signals that may originally have been considered unimportant to CNN' s systems.

- MODULAR NEURAL NETWORKS

  In this, the networks do not interfere or communicate with each other' s activities during the computation process. Consequently, big, and complex computational processes can be performed more effectively.

5.1.4. Advantages of ANN

- Having talked about the different types of Neural networks, let us look into the advantages these ANN possess:
- The information is stored on an entire network, not just a database.
- No restrictions are placed on input variables such as how they should be distributed.
- Parallel processing abilities mean that the network performs more than one job at a time, a multitasker indeed.
- It has a high fault tolerance that simply means that the corruption of one or more cells of ANN will not stop output generation.
- ANN being an intelligence system can learn from events and make decisions based on the observations.
- The ability to learn hidden relationships in the data without commanding any fixed relationship means ANN can better model highly volatile data and non-constant variance as well.
- ANN can predict the output of unseen data as its ability to generalize and infer unseen relationships is tremendous.

5.1.5. Disadvantages of ANN

- It is hardware dependent.
- It only understands numerical information so all the problems must be translated to numerical values before they are presented to the ANN.

- Lack of rules means appropriate ANN architecture can only be found by trial and error or experience.
- The inability to explain the why or how behind the solution generates a lack of trust in the network.

Have you ever wondered what forms the basis of May I know your page that meta directed you to when you were busy scrolling through or how your online signatures are verified? Nowadays, Neural networks are regulating some key sectors including finance, healthcare, and automotive. These Artificial neurons function in a way similar to the human brain. They can be used for image recognition, character recognition, and stock market predictions.

5.1.6. Applications of ANN

- Facial recognition
- Stock market prediction
- NLP, translation & language generation
- Chatbots & social media
- Aerospace & defense
- Drug discovery & healthcare development
- Signature verification & handwriting analysis
- Weather forecasting

5.1.7. Deep Learning Neural Networks

Deep learning uses neural networks to imitate how the human brain works. Thousands of interconnected artificial neurons are arranged in multiple processing layers. (Two layers are common with other machine learning systems.) The additional processing layers provide higher-level abstractions, offering better classifications and more accurate predictions. Deep learning is ideal for working with big data, voice recognition, and conversational skills.

Artificial neurons often have a weight which adjusts as the learning process proceeds. The weight increases or decreases the strength of the signal at a connection. Artificial neurons may have a threshold such that only if the aggregate signal crosses that threshold is the signal sent.

Typically, artificial neurons are organized in layers. Different layers may perform different kinds of transformations on their inputs. Signals travel from the first (input) layer to the last (output) layer, after traversing the layers multiple times.

An algorithm called feature extraction is another facet of deep learning. This facet automatically constructs meaningful "features" from the data for learning, training, and understanding.

5.2. Keras and TensorFlow

5.2.1. What is Keras?

Keras is a high-level, deep learning API developed by Google for implementing neural networks. It is written in Python and is used to make the implementation of neural networks easy. It also supports multiple backend neural network computation.

Keras is easy to learn and work with because it provides a python frontend with a high level of abstraction while having the option of multiple back-ends for computation purposes. This makes Keras slower than other deep learning frameworks, but extremely beginner friendly.

Keras allows you to switch between different back ends. The frameworks supported by Keras are:

- TensorFlow
- Theano
- PlaidML
- MXNet
- CNTK (Microsoft Cognitive Toolkit)

Out of these five frameworks, TensorFlow has adopted Keras as its official high-level API. Keras is embedded in TensorFlow and can be used to perform deep learning fast as it provides inbuilt modules for all neural network computations. At the same time, computation involving tensors, computation graphs, sessions, etc. can be custom made using the TensorFlow Core API, which gives you total flexibility and control over your application and lets you implement your ideas in a relatively short time.

5.2.2. Why Do We Need Keras?

- Keras is an API that was made to be easy to learn for people. Keras was made to be simple. It offers consistent & simple APIs, reduces the actions required to implement common code, and explains user error clearly.
- Prototyping time in Keras is less. This means that your ideas can be implemented and deployed in a shorter time. Keras also provides a variety of deployment options depending on user needs.
- Languages with a high level of abstraction and inbuilt features are slow and building custom features in then can be hard. But Keras runs on top of TensorFlow and is fast. Keras is also deeply integrated with TensorFlow, so you can create customized workflows with ease.
- The research community for Keras is vast and highly developed. The documentation and help available are far more extensive than other deep learning frameworks.
- Keras is used commercially by many companies like Netflix, Uber, Square, Yelp, etc. which have deployed products in the public domain which are built using Keras.

Apart from this, Keras has features such as:

- It runs smoothly on both CPU and GPU.
- It supports all neural network models.
- It is modular in nature, which makes it expressive, flexible, and apt for innovative research

5.2.3. How To Build a Model in Keras?

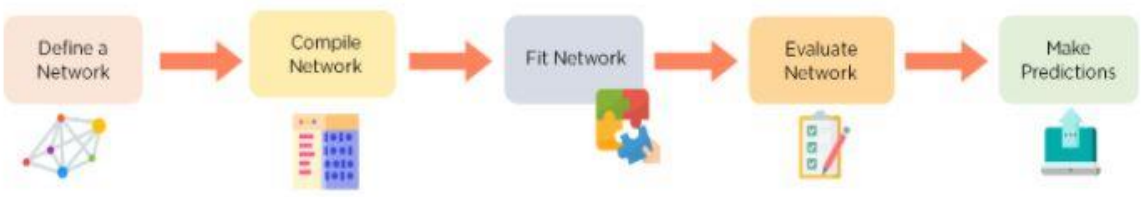

Figure 5-1 Shows the steps of how to build a model in keras

1. Define a network: In this step, you define the different layers in our model and the connections between them. Keras has two main types of models: Sequential and Functional models. You choose which type of model you want and then define the dataflow between them.

2. Compile a network: To compile code means to convert it in a form suitable for the machine to understand. In Keras, the model. Compile () method performs this function. To compile the model, we define the loss function which calculates the losses in our model, the optimizer which reduces the loss, and the metrics which is used to find the accuracy of our model.

3. Fit the network: Using this, we fit our model to our data after compiling. This is used to train the model on our data.

4. Evaluate the network: After fitting our model, we need to evaluate the error in our model.

5. Make Predictions: We use model. Predict () to make predictions using our model on new data.

5.2.4. Applications of Keras

- Keras is used for creating deep models which can be productized on smartphones.
- Keras is also used for distributed training of deep learning models.
- Keras is used by companies such as Netflix, Yelp, Uber, etc.

Keras is also extensively used in deep learning competitions to create and deploy working models, which are fast in a short amount of time.

5.2.5. What is TensorFlow?

TensorFlow is an open-source end-to-end platform for creating Machine Learning applications. It is a symbolic math library that uses dataflow and differentiable programming to perform various tasks focused on training and inference of deep neural networks. It allows developers to create machine learning applications using various tools, libraries, and community resources.

Currently, the most famous deep learning library in the world is Google' s TensorFlow. Google product uses machine learning in all of its products to improve the search engine, translation, image captioning or recommendations.

5.2.6. Why is TensorFlow Popular?

TensorFlow is the best library of all because it is built to be accessible for everyone. TensorFlow library incorporates different API to build at scale deep learning architecture like CNN or RNN. TensorFlow is based on graph computation; it allows the developer to visualize the construction of the neural network with Tensorboad. This tool is helpful to debug the program. Finally, TensorFlow is built to be deployed at scale. It runs on CPU and GPU.

TensorFlow attracts the largest popularity on GitHub compared to the other deep learning framework.

5.2.7. TensorFlow Algorithms

Below are the algorithms supported by TensorFlow:

Currently, TensorFlow 1.10 has a built-in API for:

- Linear regression: tf. estimator. LinearRegressor
- Classification:tf. estimator. LinearClassifier
- Deep learning classification: tf. estimator. DNNClassifier
- Deep learning wipe and deep: tf. estimator. DNNLinearCombinedClassifier
- Booster tree regression: tf. estimator. BoostedTreesRegressor
- Boosted tree classification: tf. estimator. BoostedTreesClassifier

5.3. Neural Network Activation Function

5.3.1. Definition

An Activation Function decides whether a neuron should be activated or not. This means that it will decide whether the neuron' s input to the network is important or not in the process of prediction using simpler mathematical operations.

The role of the Activation Function is to derive output from a set of input values fed to a node (or a layer).

if we compare the neural network to our brain, a node is a replica of a neuron that receives a set of input signals—external stimuli.

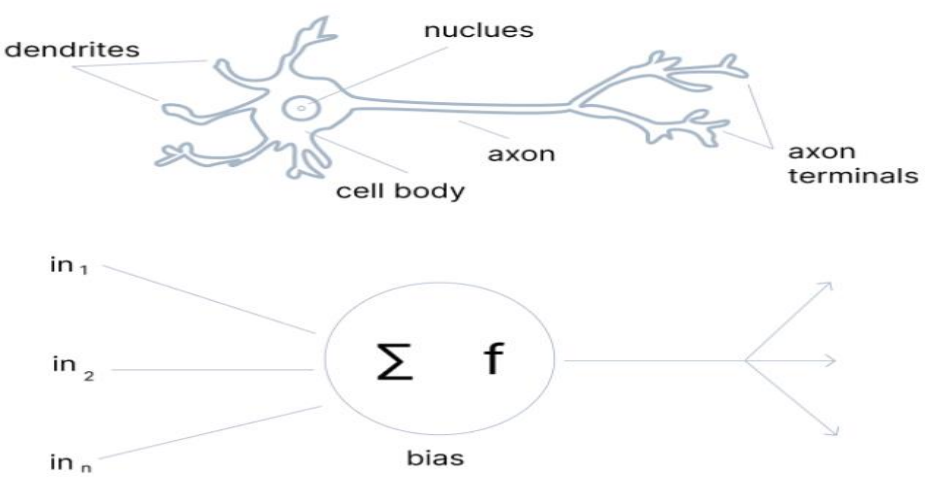

Figure 5-2 Structure of neural network

Depending on the nature and intensity of these input signals, the brain processes them and decides whether the neuron should be activated ("fired") or not.

In deep learning, this is also the role of the Activation Function—that' s why it' s often referred to as a *Transfer Function* in Artificial Neural Network.

The primary role of the Activation Function is to transform the summed weighted input from the node into an output value to be fed to the next hidden layer or as output.

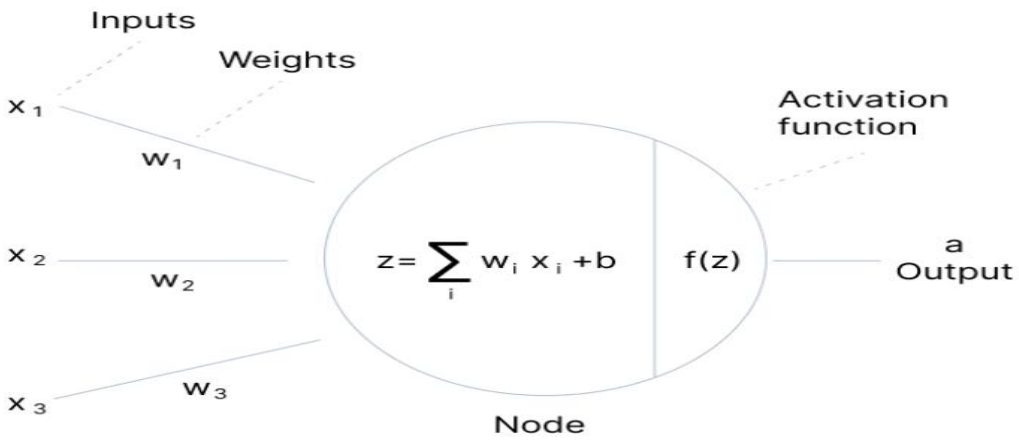

Figure 5-3 The use of the activation function

5.3.2. Elements of a Neural Network Architecture

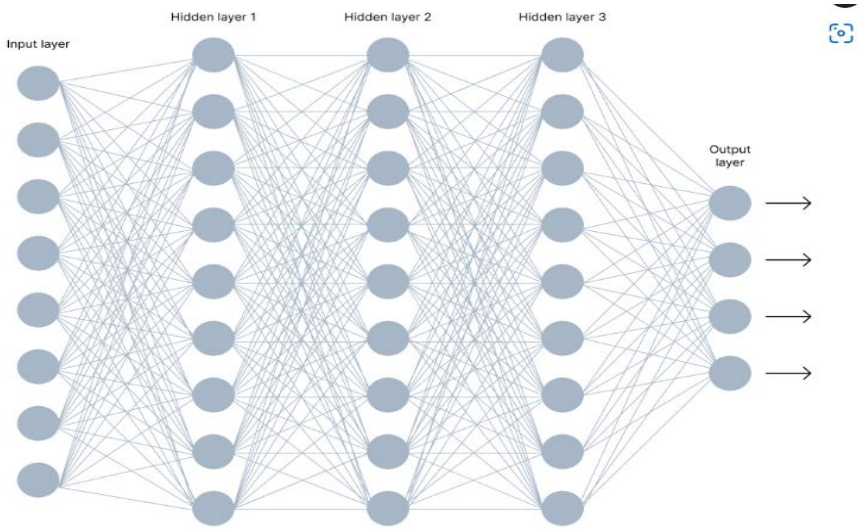

Figure 5-4 Detailed neural network

In the image above, you can see a neural network made of interconnected neurons. Each of them is characterized by its **weight**, **bias**, and **activation function**

Input Layer

The input layer takes raw input from the domain. No computation is performed at this layer. Nodes here just pass on the information (features) to the hidden layer.

Hidden Layer

As the name suggests, the nodes of this layer are not exposed. They provide an abstraction to the neural network.

The hidden layer performs all kinds of computation on the features entered through the input layer and transfers the result to the output layer.

Output Layer

It is the final layer of the network that brings the information learned through the hidden layer and delivers the final value as a result.

Note*:* All hidden layers usually use the same activation function. However, the output layer will typically use a different activation function from the hidden layers. The choice depends on the goal or type of prediction made by the model.

5.3.3. Why do Neural Networks Need an Activation Function?

Activation functions introduce an additional step at each layer during the forward propagation, but its computation is worth it. Here is why—

Let us suppose we have a neural network working *without* the activation functions.

In that case, every neuron will only be performing a linear transformation on the inputs using the weights and biases. It is because it does not matter how many hidden layers we attach in the neural network; all layers will behave in the same way because the composition of two linear functions is a linear function itself.

Although the neural network becomes simpler, learning any complex task is impossible, and our model would be just a linear regression model.

5.3.4. Types of Neural Networks Activation Functions

There are three types of neural networks activation functions

- Binary Step Function

Binary step function depends on a threshold value that decides whether a neuron should be activated or not.

The input fed to the activation function is compared to a certain threshold; if the input is greater than it, then the neuron is activated, else it is deactivated, meaning that its output is not passed on to the next hidden layer.

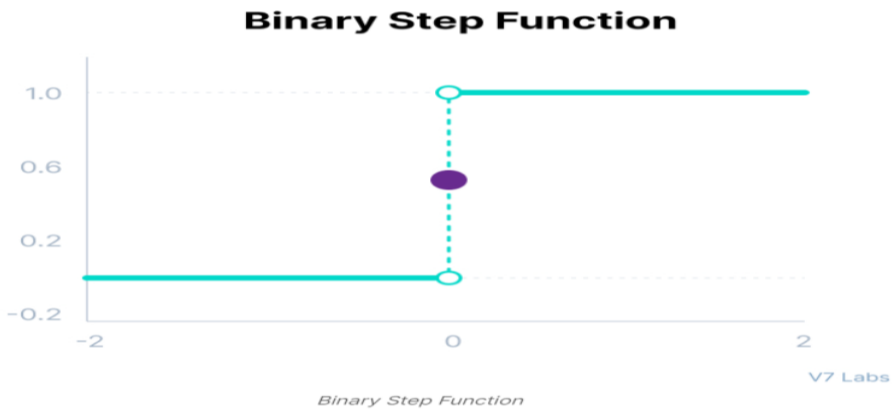

Figure 5-5 Binary Step Function

- Linear Activation Function

The linear activation function, also known as "no activation," or "identity function" (multiplied x1.0), is where the activation is proportional to the input.

The function does not do anything to the weighted sum of the input, it simply spits out the value it was given.

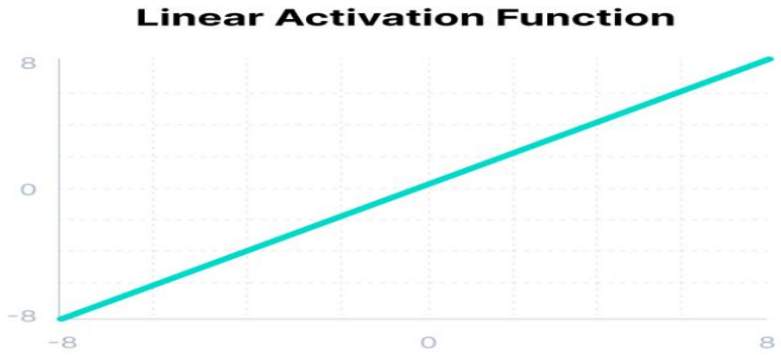

Figure 5-6 Linear Activation Function

- Non-Linear Activation Functions

The linear activation function shown above is simply a linear regression model.

Because of its limited power, this does not allow the model to create complex mappings between the network' s inputs and outputs.

Non-linear activation functions solve the following limitations of linear activation functions:

- They allow backpropagation because now the derivative function would be related to the input, and it is possible to go back and understand which weights in the input neurons can provide a better prediction.

- They allow the stacking of multiple layers of neurons as the output would now be a non-linear combination of input passed through multiple layers. Any output can be represented as a functional computation in a neural network.

There are 10 Non-Linear Neural Networks Activation Functions:

1. **Sigmoid / Logistic Activation Function**

This function takes any real value as input and outputs values in the range of 0 to 1.

The larger the input (more positive), the closer the output value will be to 1.0, whereas the smaller the input (more negative), the closer the output will be to 0.0, as shown below.

Mathematically it can be represented as:

$$f(x) = \frac{1}{1 + e^{-x}}$$

2. **Tanh Function (Hyperbolic Tangent)**

Tanh function is very similar to the sigmoid/logistic activation function, and even has the same S-shape with the difference in output range of -1 to 1. In Tanh, the larger the input (more positive), the closer the output value will be to 1.0, whereas the smaller the input (more negative), the closer the output will be to -1.0.

$$f(x) = \frac{\left(e^{x} - e^{-x}\right)}{\left(e^{x} + e^{-x}\right)}$$

3. ReLU Function

ReLU stands for Rectified Linear Unit.

The main catch here is that the ReLU function does not activate all the neurons at the same time.

The neurons will only be deactivated if the output of the linear transformation is less than 0.

$$f(x) = max\,(0, x)$$

4. Leaky ReLU Function

Leaky ReLU is an improved version of ReLU function to solve the Dying ReLU problem as it has a small positive slope in the negative area.

$$f(x) = max\,(0.1x, x)$$

5. Parametric ReLU Function

Parametric ReLU is another variant of ReLU that aims to solve the problem of gradient' s becoming zero for the left half of the axis.

$$f(x) = max\,(ax, x)$$

6. Exponential Linear Units (ELUs) Function

ELU uses a log curve to define the negative values unlike the leaky ReLU and Parametric ReLU functions with a straight line.

$$\begin{cases} x & for\ x \geqslant 0 \\ \alpha(e^x - 1) & for\ x < 0 \end{cases}$$

7. SoftMax Function

Mathematically it can be represented as:

$$\text{softmax}(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$$

8. Swish

Swish consistently matches or outperforms ReLU activation function on deep networks applied to various challenging domains such as image classification, machine translation etc.

$$f(x) = x*sigmoid(x)$$

9. Gaussian Error Linear Unit (GELU)

The Gaussian Error Linear Unit (GELU) activation function is compatible with BERT, ROBERTa, ALBERT, and other top NLP models. This activation function is motivated by combining properties from dropout, zoneout, and ReLUs.

Mathematically it can be represented as:

$$f(x) = xP(X \leq x) = x\Phi(x)$$
$$= 0.5x\left(1 + tanh\left[\sqrt{2/\pi}\left(x + 0.044715x^3\right)\right]\right)$$

10. Scaled Exponential Linear Unit (SELU)

SELU was defined in self-normalizing networks and takes care of internal normalization which means each layer preserves the mean and variance from the previous layers. SELU enables this normalization by adjusting the mean and variance

Mathematically it can be represented as:

$$f(\alpha, x) = \lambda\begin{cases}\alpha\left(e^x - 1\right) & for\ x < 0\\ x & for\ x \geqslant 0\end{cases}$$

Figure 5-7 Neural Network Non-Linear Activation Functions

5.3.5. Rules for Choosing the Activation Function

a few rules for choosing the activation function for your output layer based on the type of prediction problem that you are solving:

6. **Regression** - Linear Activation Function
7. **Binary Classification**—Sigmoid/Logistic Activation Function
8. **Multiclass Classification**—Softmax
9. **Multilabel Classification**—Sigmoid

The activation function used in hidden layers is typically chosen based on the type of neural network architecture.

5. **Convolutional Neural Network (CNN)**: ReLU activation function.
6. **Recurrent Neural Network**: Tanh and/or Sigmoid activation function

6. Chapter 6: Using Neural Network to Solve the Inverse Problem

To do our inverse problem we used neural network. We used server to connect with forward which made in C++ to get data.

In this chapter we will talk in detail about our code. We will do this with steps in three phases.

6.1. First phase: connection between two codes

In our project we use C++ to visualize and produce the data of heart and the body and python to use neural network by getting data which C++ produces to train NN then using this model to predict outer data and return it to C++ to visualize it. We used socket library from python and do two libraries (serializer and client) to use them in connection between python and C++

6.2. Second phase: Training data

This phase divided to steps:

1- In this step we get data from C++ and format it in csv file to be easy to be processed.

2- **Dividing data to input and output:** In this step we divide data as we want first **70** column in csv file as input (torso) and the last **30** column as output (heart)
We take the data from forward model with **0.0001**-time step.

70 probes on torso

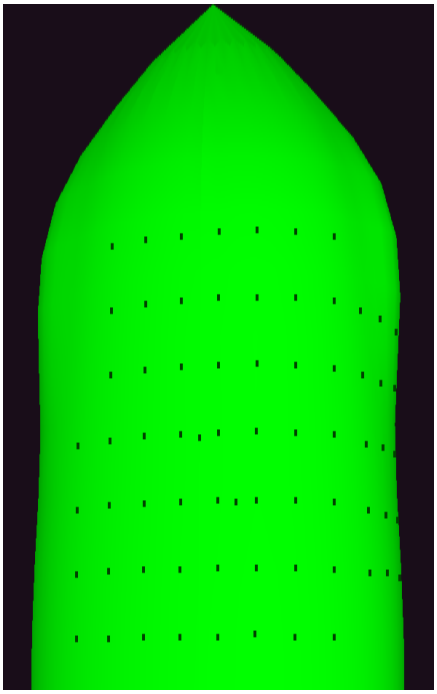

Figure 6-1a Distribution of probes on the front view of the torso

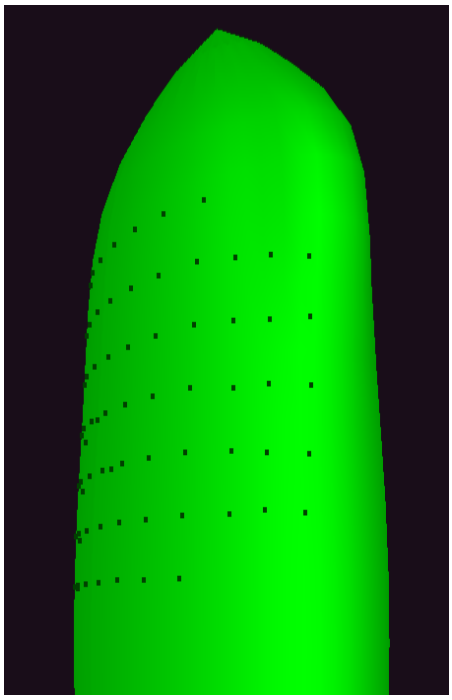

Figure 6-1b Distribution of probes on the side view of the torso

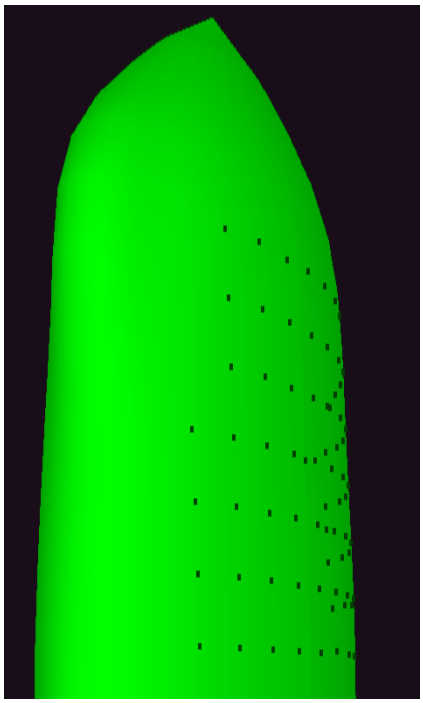

Figure 6-1c Distribution of probes on the other side view of the torso

30 probes on heart

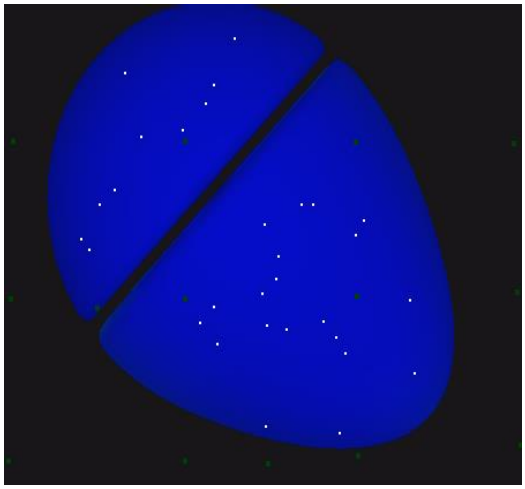

Figure 6-2a Probes of the front of the heart

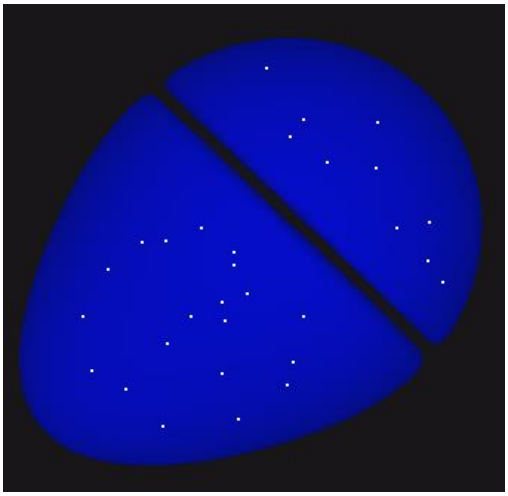

Figure 6-2b Probes of the back of the heart

Some random probes of ECG

Figure 6-3a ECG of probe no. 16

Figure 6-3b ECG of probe no. 36

Figure 6-3c ECG of probe no. 54

Figure 6-3d ECG of probe no. 8

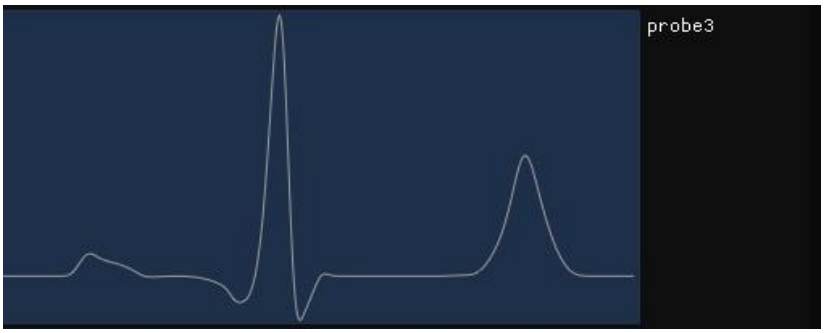

Figure 6-3e ECG of probe no. 3

As we see more than you are near standard ECG probes the wave will be near to knowing shape.

3- **Normalization:** In this step we normalize data by using tensorflow library with function MinMaxScaler() so fit with activation functions need normalized data .

4- **Splitting data:** We split all data (input and output data) to train, validation and test sets. we divide data to 30% of it test and validation then take it and divided to 50% validation and 50% test.

Figure 6-4 Split Data

5- **Train the neural network:**

We create a Sequential model and add layers one at a time until we are happy with our network architecture. The first thing to get right is to ensure the input and output layers have the right number of features. This can be specified when creating the first layer with the input shape argument and setting it to (70,) for presenting the 70 input torso probes variables as a vector and 30 as output (heart probes). We can specify the number of neurons or nodes in the layer as the first argument and specify the activation function using the activation argument. We will use ReLU on all hidden layers and sigmoid in the output layer.

ReLU Function

ReLU stands for Rectified Linear Unit.

The main catch here is that the ReLU function does not activate all the neurons at the same time.

The neurons will only be deactivated if the output of the linear transformation is less than 0.

$$f(x) = max\ (0, x)$$

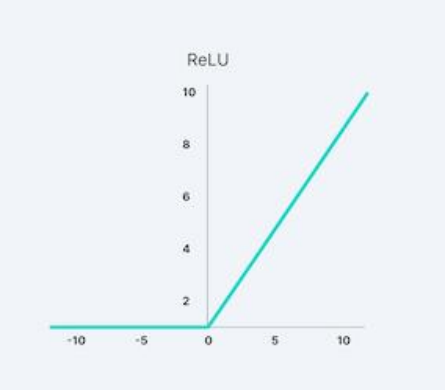

Figure 6-5 Relu Function

After we choose the activation function and the weights of each layer, we compile the model and choose the ***optimizer***. The compile method configures the model' s learning process. First of all, the best important one is the optimizer because the core of any neural network is the optimizer. In order to

determine the appropriate weights for prediction, each neural network optimizes a loss function. There are several different types of optimizers, each with a somewhat different technique to identifying the optimal weights. Optimizers are procedures or methodologies for minimizing an error function as well as increasing production efficiency. Optimizers are computational models that are built on the learnable parameters of a model, such as Weights and Biases. Optimizers assist in determining how to adjust the weights and learning rate of a neural network in order to minimize losses. In layman's terms, optimizers tinker with the weights to shape and mould your model into the most accurate form possible. The loss function serves as a road map for the optimizer, indicating whether it is travelling in the correct or wrong path.

Types of Optimizers:

1- Gradient Descent
2- Stochastic Gradient Descent.
3- Mini-Batch Gradient Descent.
4- Momentum Based Gradient Descent.
5- Nesterov Accelerated Gradient
6- Adagrad.
7- RMSProp
8- Adam

Losses in Neural Network

Loss functions are required while compiling a model. This loss function would be optimized by the optimizer, which was also specified as a parameter in the compilation procedure. Probabilistic losses, regression losses, and hinge losses are the three types of losses.

One of the most key aspects of Neural Networks is the Loss Function. Loss is nothing more than a Neural Net prediction error. Loss Function is the way for computing the loss.

To put it another way, the gradients are calculated using the Loss. Gradients are also utilized to update the Neural Net's weights.

Keras and TensorFlow have a variety of loss functions built-in for varied goals.

It's a method of determining how effectively your algorithm models the data. Your loss function will produce a greater value if your forecasts are completely wrong. If they're decent, it'll give you a lower number. Your loss function will inform you if you're making progress as you tweak parts of your algorithm to try to enhance your model.

A matrix is a function that is used to judge the performance of your model.

Metrics functions are similar to loss functions, except that the results from evaluating a matrix are not used when training the model. Note that you may use any loss function as a matrix.

Available metrics

1- Accuracy metrics
2- Probabilistic metrics
3- Regression metrics
4- Classification metrics based on True/False positives & negatives
5- Image segmentation metrics
6- Hinge metrics for "maximum-margin" classification

Evaluation Metrics

To assess its performance on test data, every Machine Learning model use metrics. Metrics are similar to Loss Function; except they are used to monitor test data. Accuracy, Binary Accuracy, Categorical Accuracy, and other forms of accuracy measures exist. Probabilistic measures such as binary cross-entropy, categorical cross-entropy, and others are also provided.

It also consists of regression metrics such as Root Mean Squared Error, Means Absolute Error, Means Squared Error, etc....

Different kinds of evaluation metrics available are as follows

1- Classification Accuracy
2- Confusion Matrix
3- Area under Curve
4- Logarithmic Loss
5- Mean Squared Error
6- F1 Score
7- Mean Absolute Error

In our code, we choose the optimizer *"adam"* as we define it is the suitable for our data and get best accuracy also, we choose *mean_squared_error* method to evaluate the losses in prediction because we deduce that this method produces the best accuracy finally we choose in metric *accuracy* class.

Next step in this phase that we fit model and divide it to batches and epochs

To evaluate the accuracy and losses in train and validation data. in our code we divide data to *115 epochs and 16 batches* this produce best results.

All above we in the train model step which the most important step in all code if we do it well all the code will be easier. now go to the next step.

6 – plot the accuracy and losses: now after evaluating the accuracy and losses we will plot them in train and validation data by using matplotlib library.

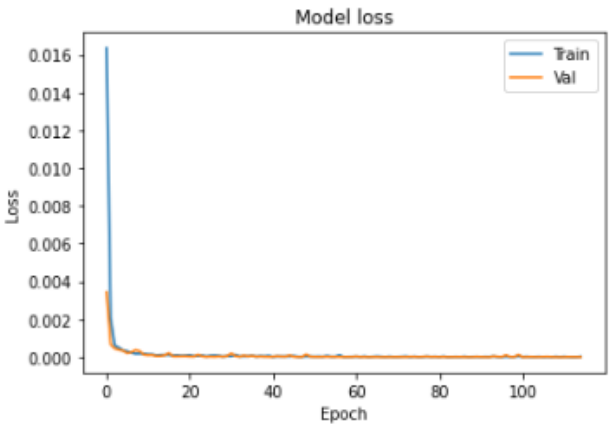

Figure 6-6 Model Loss

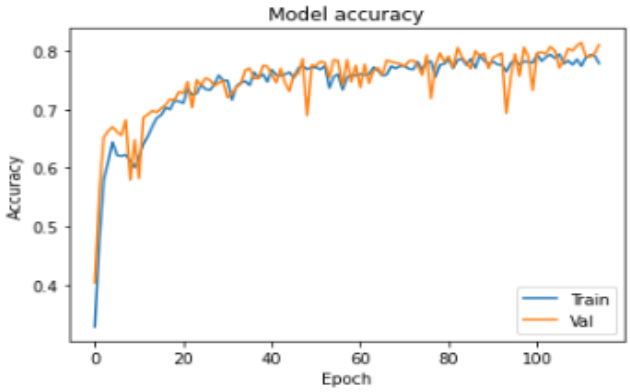

Figure 6-7 Model Accuracy

Test set has an accuracy equal to 81.16%

7 - Last step we will save the accuracy and weights:

Above we use training and validation data to evaluate accuracy and loss now we use the evaluation method to evaluate test data. we should save weights in file to save weights which produce best accuracy to return when we predict any other data.

6.3. Third phase: prediction

From now we will talk about predicting data . also we will do this in steps .

1- Getting data and normalizing it

in this step we will get data from a csv file (external data) to predict heart potentials then as before we will normalize data to use it .

2- Prediction

we predict the output from the NN model which we made.

3- Visualizion

Finally we send the output to C++ to visualize it .

6.4. Result and conclusion

For normal heart

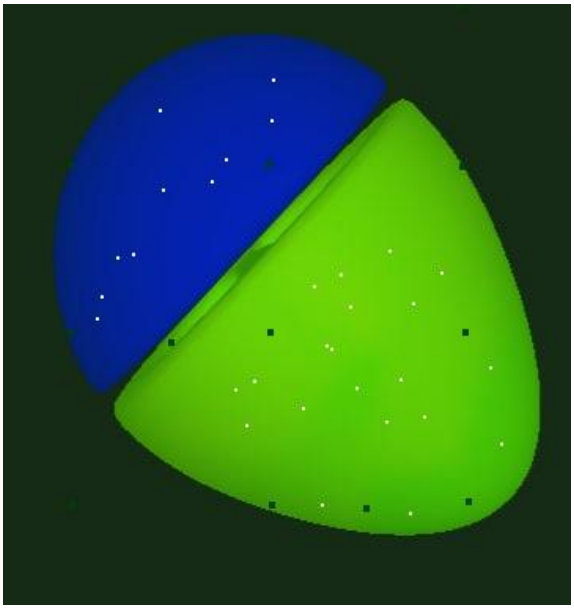

Figure 6-8a Propagation wave of the heart after prediction

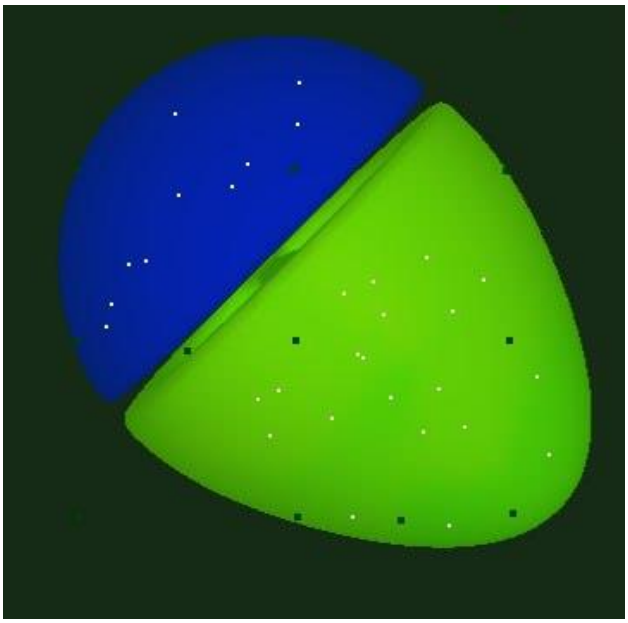

Figure 6-8b Propagation wave of the heart before prediction

Actual potential of heart at random points after prediction

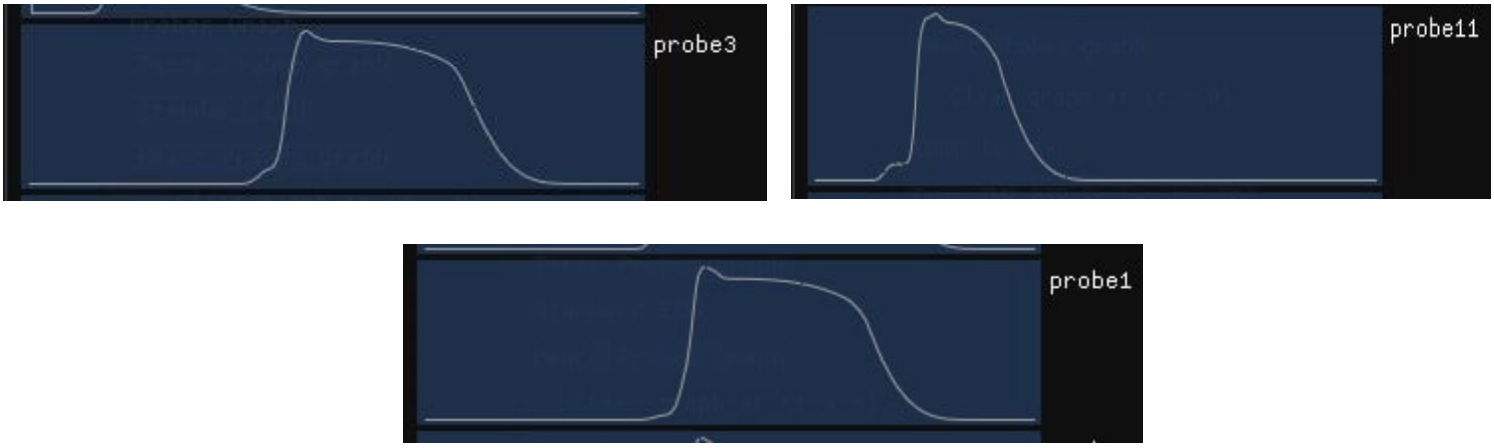

Figure 6-9 Action potential of random heart probes

Abnormal heart

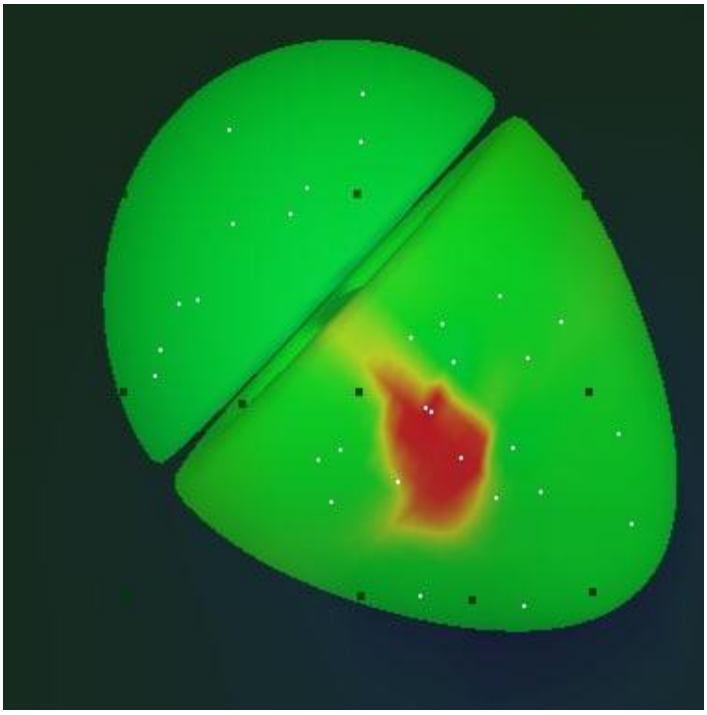

Figure 6-10a Propagation wave of the infracted heart after prediction

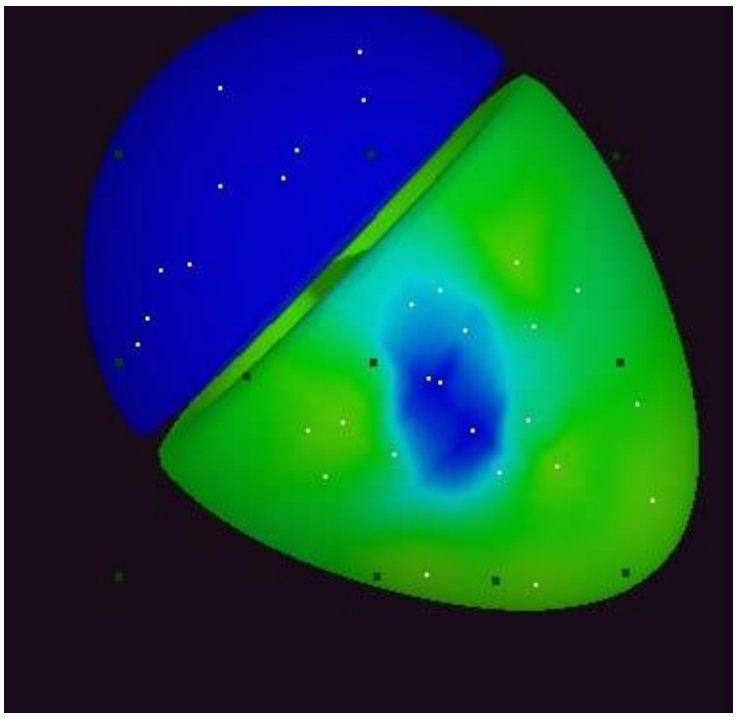

Figure 6-10b Propagation wave of the infracted heart before prediction

Actual potential of heart at random points after prediction

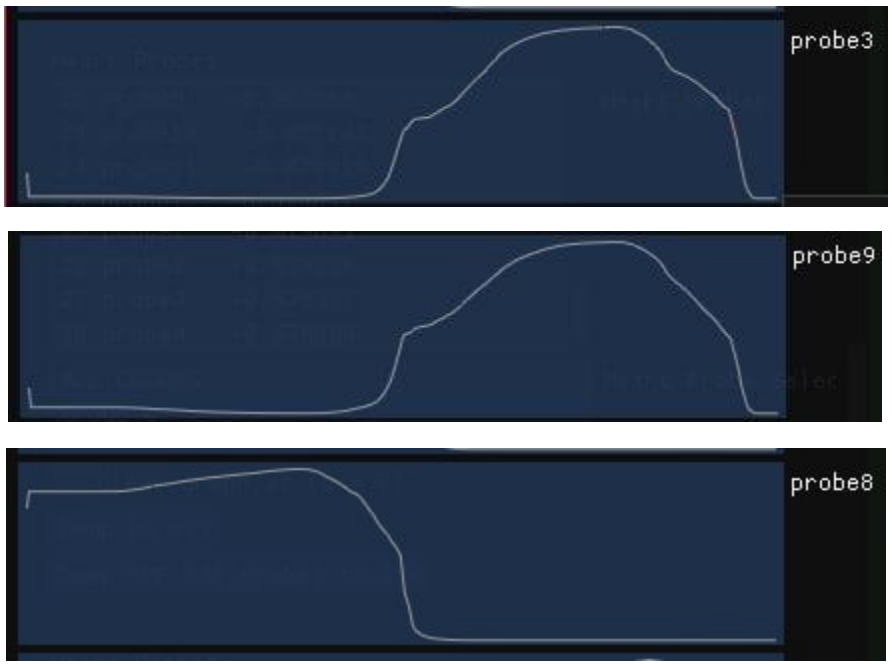

Figure 6-11 Action potential of random probes after prediction

7. Chapter 7: Hologram

7.1. What is hologram?

Hologram technology is one of the three-dimensional imaging methods that enable the formation of an image with its three dimensions using light **diffraction and interference**, this technology gives us a broader picture that is close to reality and so more information about the body that it represents in a 3D model. A Hologram (that is pronounced as HOL-O-Gram), is a special type of 3D image which can be seen with the naked eye the term ' Hologram' comes from Dennis Gabor who coined it from the Greek word 'holos ' which means "all" ,and with "gram " which means " message ", the technology of creating and presenting these holograms using different methods and tools is called Hologram technology (Slinger et al. 46–53: 38). Holography is a photographic technique that records the light **scattered** from an object and then presents it as a three-dimensional image, to formulate it briefly the definition of holography could be the science and practice of making Holograms. A Hologram is a truly three-dimensional free-standing image that is created with a photographic projection of light, it differs from other technologies that present a three-dimensional object on a two-dimensional computer display or using special wearable glasses, Holograms could be seen with the naked eye because it is a physical structure that is created using laser beams or light diffraction.

The term hologram itself was composed of Greek words holo = whole and gramma = message.

7.2. How it is work:

Holography is a technique that enables a light field (which is generally the result of a light source scattered off objects) to be recorded and later reconstructed when the original light field is no longer present, due to the absence of the original objects. Section 1 Holography can be thought of as somewhat similar to sound recording, whereby a sound field created by vibrating matter like musical instruments or vocal cords, is encoded in such a way that it can be reproduced later, without the presence of the original vibrating matter. However, it is even more like Amnionic sound recording in which any listening angle of a sound field can be reproduced in the reproduction.

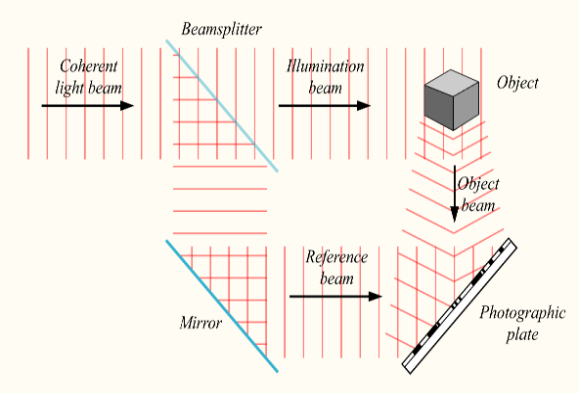

Figure 7-1 Hologram Working Principle

7.3. Image properties:

The image is a 3D, really existed, and float in midair. The observer can see the image directly without any additional devices like the 3D glasses or smartphone screen.

Figure 7-2 A holographic image

7.4. Image creation:

There are many types fall under the 3D hologram technology and each type differs in the method of image creation and projection. The image creation is done by two main methods.

the first is image creation using interferometer and transmission. This method includes two categories of hologram creating process, the Reflection and Transmission Hologram.

- **Reflecting Hologram:** the laser beam splits into an object beam and a reference beam. Two mirrors placed in front of the object beam and the reference beam to reflect the beam and form a 3D hologram on the photographic plate.

- **Transmission Hologram:** the laser splits to form an object beam and a reference beam. Then the illuminated beam from object and reference reflects into photographic plate to form a 3D hologram.

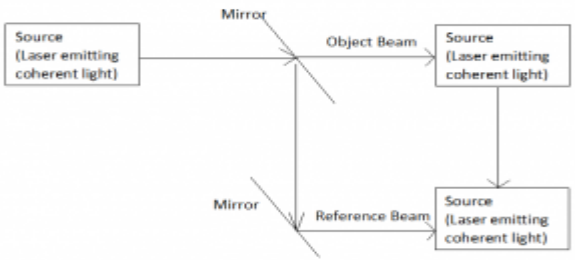

Figure 7-3 Transmission Hologram Principle

Another method to create the 3D hologram is a computer-generated hologram (CGH).

CGH gives us an advanced method by creating a 3D visual representation for the object and calculating the holographic fringe pattern then use it to set the optical properties via a spatial light modulator (SLM) such as liquid crystal micro display. The laser is required in CGH to display the 3D hologram because of its spectral purity.

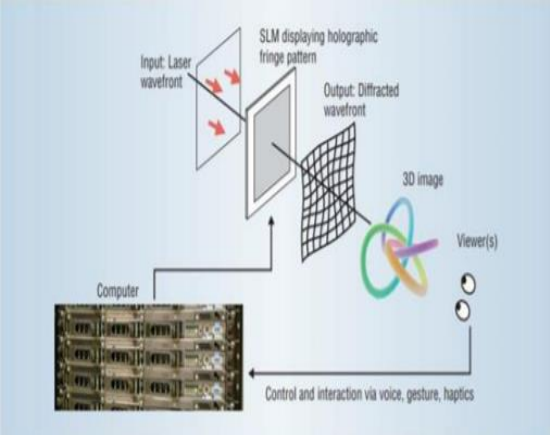

Figure7-4 Computer Generated Hologram (CGH)

7.5. Hologram technology requirements:

The specification of requirements is very important in every system to be sure that it will reach the wanted goals, these needs differ from system to system and also according to different technologies used, there are some technologies that have special requirements needed to be used even when used with a different type of systems and situations and it will not be present without it, other requirements may be added or changed as systems changes.

7.5.1. User requirements

Humans are visual creatures, and they have a 3D perception when they look at an object, the retina in each eye forms a two-dimensional image of our surroundings and the brain processes these two images into a 3D visual experience, as a result, it is important for them to get a semi-real image when communicating visually, 3D objects offer users a larger perception and so more information to get. When the technology can give the user the possibility of a complete view of a not actually present object, it will be considered as a very useful technology that results in users satisfy.

7.5.2. Hardware requirements

The great development witnessed by these two technologies led to a big difference in the devices used over time. Large and multiple devices were used in the past, or now, the development of technology led to the development of the quality of the devices used, at the present time advanced devices of different sizes and features are used.

Hologram Technology Hardware requirements: Many devices involved in using the Hologram technology, starting from the cameras used to take pictures of the object being displayed to the projector that will display it. Some of the most important hardware needed in Hologram technology are Cameras, Computers \ Smartphone \ Remote control, and light source generation device.

Hologram projectors: Hologram technology needs a special projector to display the Holograms, and as the Hologram is an existing image the Hologram projector is responsible for creating it in the surround so it can be seen by the naked eye. There are different types of projectors to display the 3D image, some of it is very simple using general light concepts to view the Holograms and others are so complicated and needed several steps and devices to generate the images.

There are several types of Hologram projectors that display the holograms using different methods, including the hologram projectors that create Holograms that are scalable to interact directly, also, some other types are used mainly in theatrical performances, and the type which creates a hologram using different media such as **smoke, water, etc**. Below are the most used and known types. Figure 7.5 shows the **hologram fan** projector and in Figure 7.6 **a pyramid hologram** projector is displayed and finally Figure 7.7 displays **the laser hologram projector** which came in many forms.

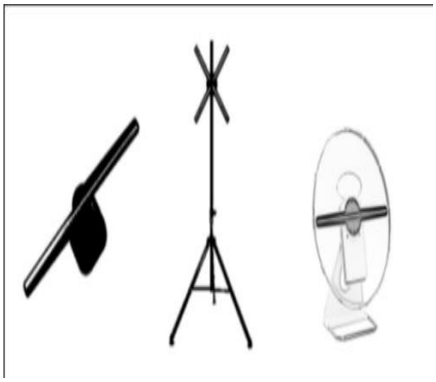

Figure 7-5 Hologram Fan

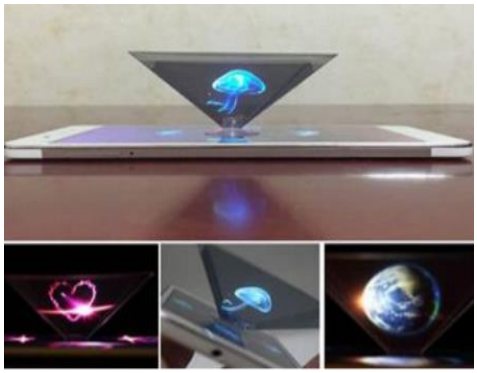

Figure 7-7 Pyramid Hologram

Figure 7-6 Laser Hologram

7.5.3. Software requirements

Hologram technology, on the other hand, is very flexible in the type of software needed, as the needed hardware differs and result in more technologies that could be used, and the most needed software to applying Hologram technology are:

Operating system, Programming language, 3D objects crating and photo editing software, Hologram device programming software and Hologram projector controlling software.

7.6. Hologram technology types:

Hologram technology has various types, which are classified differently based on the way of recording, recording and display environment, and reconstructing way, and this is an advantage to favor the hologram to be used in many areas and different purposes. Some 3D Hologram types, which are widely used are:

Reflection Holograms, Transmission Holograms, Laser plasma Hologram, No-logram, Smoke Hologram, Underwater Hologram, HoloLamp, Fairy light Hologram, Physical Holograms, 360°Hologram and Stereogram.

The Hologram types also classified depending on the projector device, and we are going to list three main types:

- **Hologram Pyramid:** it uses the concept of light interference and reflection.
- **Laser Hologram**: a laser of different types forms the image.
- **Hologram fan:** This type of projector display Holograms using small propellers merged with high tech RGB lights that can switch colors in milliseconds.

Laser Holography:

In laser holography, the hologram is recorded using a source of laser light, which is very pure in its color and orderly in its composition. Various setups may be used, and several types of holograms can be made, but all involve the interaction of light coming from different directions and producing a microscopic interference pattern which a plate, film, or other medium photographically records.

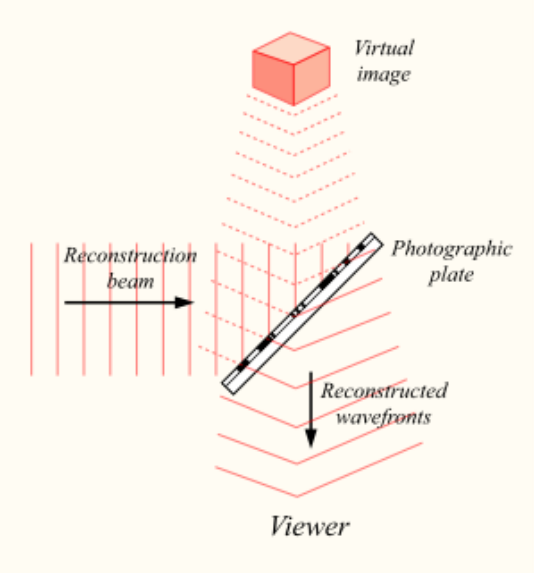

Figure 7-8 Laser Holography Principle

In one common arrangement, the laser beam is split into two, one known as the object beam and the other as the reference beam. The object beam is expanded by passing it through a lens and used to illuminate the subject. The recording medium is located where this light, after being reflected or scattered by the subject, will strike it. The edges of the medium will ultimately serve as a window through which the subject is seen, so its location is chosen with that in mind. The reference beam is expanded and made to shine directly on the medium, where it interacts with the light coming from the subject to create the desired interference pattern.

7.7. Holographic data storage

Holographic data storage is a potential technology in high capacity data storage. While magnetic and optical data storage devices rely on individual bits being stored as distinct magnetic or optical changes on the surface of the recording medium, holographic data storage records information throughout the volume of the medium and is capable of recording multiple images in the same area utilizing light at different angles.

Additionally, whereas magnetic and optical data storage records information a bit at a time in a linear fashion, holographic storage is capable of recording and reading millions of bits in parallel, enabling data transfer rates greater than those attained by traditional optical storage.

7.7.1. Recording Data

Holographic data storage contains information using an optical interference pattern within a thick, photosensitive optical material. Light from a single laser beam is divided into two, or more, separate optical patterns of dark and light pixels. By adjusting the reference beam angle, wavelength, or media position, a multitude of holograms (theoretically, several thousands) can be stored on a single volume.

7.7.2. Reading Data

The stored data is read through the reproduction of the same reference beam used to create the hologram. The reference beam's light is focused on the photosensitive material, illuminating the appropriate interference pattern, the light diffracts on the interference pattern, and projects the pattern onto a detector. The detector can read the data in parallel, over one million bits.

7.8. Hologram technology usage

Hologram technology is used in different fields, their useful result enables them to bring about development and improvement in many areas of life to make human lives easier and more developed. The two technologies aim to produce the same result, and mostly they are able to be applied and used in the same life fields, but there are situations when the technology will not be able to give the proper result or make users satisfied and these situations differ according to the technology's capabilities and properties. Some common field in which both technologies could be applied to are Education, Advertising, Museums, Military Weather and maps, Astronomy, Training, Entertainment, Advising.

7.8.1. Medicine and Imaging

Hologram technology is on its way to revolutionizing medicine. It has the capability to produce a full- color 3D hologram of the human body. Students and doctors can visualize the three-dimensional images of complex organs like the **brain, heart, liver, lungs, nerves, and muscles**. This technology can also help in surgical pre-planning. Before real surgery, the surgeon can fully visualize the entire course of the operation and thereby increase the chances of a successful outcome on patients. Digital holographic microscopy makes it possible to perform cell counting and analysis of subcellular motion deep in living tissue. It also supports simultaneous imaging at different depths.

Recently, Royal Philips and RealView Imaging reported a clinical study demonstrating the feasibility of using an innovative live 3D holographic visualization and interaction technology to guide minimally invasive structural heart disease procedures. The study involved 8 patients and was conducted in collaboration with the Schneider Children' s Medical Center in Petach Tikva, Israel. RealView' s visualization technology was used to display interactive, real-time 3D holographic images acquired by Philips' interventional X-ray and cardiac ultrasound systems. Consequently, in addition to viewing the patient' s heart on a 2D screen, interventional cardiologists were able to view detailed dynamic 3D holographic images of the heart 'floating in free space' during an interventional structural heart disease procedure, without using special eyewear. The physicians were also able to manipulate the projected 3D heart structures by literally touching the holographic volumes in front of them. The study demonstrated the potential of the technology to enhance the context and guidance of structural heart repairs. The holographic projections enabled not only to intuitively understand and interrogate the 3D spatial anatomy of the patient' s heart, but also to navigate and appreciate the device-tissue interaction during the procedure.

There are different companies producing true hologram systems focused on the medical sector.

- EchoPixel, from the USA, produces a software capable of converting 2D images into stereoscopic 3D images. With this system, professionals can 'cut' virtual tissue, organs, or other parts of the body at different angles, enabling surgeons and image specialists to create endless number of cross sections.
- 3DforScience from Spain, Ovizio Imaging systems from Belgium, Holoxica Ltd., from UK, Lyncee Tec and NanoLive from Switzerland or RealView Imaging from Israel, are other companies that focues in the global medical holography markets.

7.9. Captions of our model on the hologram

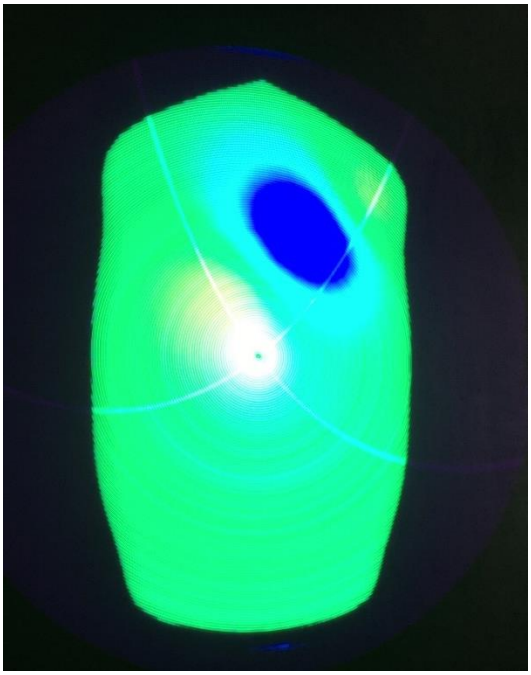

Figure 7-9 Caption of the torso simulation

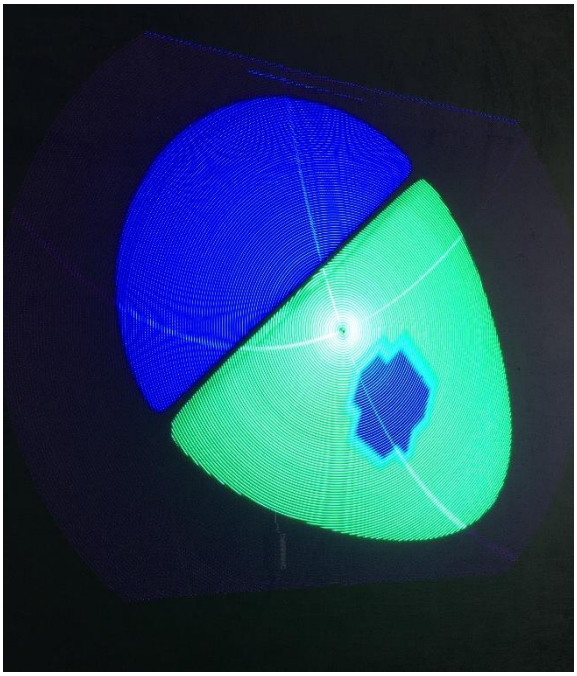

Figure 7-10 Caption of the heart simulation

References

1. Comprehensive Electrocardiology 2nd Edition; Peter W. Macfarlane, A. van Oosterom, Olle Pahlm, Paul Kligfield, Michiel Janse, John Camm,2010
2. John Hopkins Medicine; Anatomy and Function of the Heart's Electrical System.
3. Cardiovascular Physiology Concepts New Third Edition; Wolters Kluwer, 2021
4. Reconstructing the heart depolarization pattern from body surface potential using artificial neural network; N. A. Mabrouk, A. M. Khalifa, A. A. A. Nasser, and Moustafa H. Aly
5. D. B. Geselowitz, "On the theory of the electrocardiogram," in Proceedings of the IEEE, vol. 77, no. 6, pp. 857-876, June 1989, doi: 10.1109/5.29327.
6. Geselowitz DB. On bioelectric potentials in an inhomogeneous volume conductor. Biophys J. 1967 Jan;7(1):1-11. doi: 10.1016/S0006-3495(67)86571-8. Epub 2008 Dec 31. PMID: 19210978; PMCID: PMC1368053.
7. Geselowitz DB, Miller WT 3rd. A bidomain model for anisotropic cardiac muscle. Ann Biomed Eng. 1983;11(3-4):191-206. doi: 10.1007/BF02363286. PMID: 6670784.
8. Barr RC, Ramsey M 3rd, Spach MS. Relating epicardial to body surface potential distributions by means of transfer coefficients based on geometry measurements. IEEE Trans Biomed Eng. 1977 Jan;24(1):1-11. doi: 10.1109/TBME.1977.326201. PMID: 832882.
9. Tikhonov A. and V. Arsenin, Solution of Ill-posed Problems. Winston, Washington, DC, 1977.
10. Greensite F., Well-posed formulation of the inverse problem of electrocardiography. Ann. Biomed. Eng., 1944;22: 172-183
11. Yamashita Y., Theoretical studies on the inverse problem in electrocardiography and the uniqueness of the solution. IEEE Trans, Biomed. Eng., 1982;29: 719–725
12. Brooks D., G. Ahmad, and R. MacLeod, Multiply constrained inverse electrocardiology: Combining temporal, multiple spatial, and iterative regularization, in Proceedings of the IEEE Engineering in Medicine and Biology Society 16th Annual International Conference. IEEE Computer Society, 1994, pp. 137–138
13. Ripley, Brian D. (1996) Pattern Recognition and Neural Networks, Cambridge.
14. Bishop, C.M. (1995) Neural Networks for Pattern Recognition, Oxford: Oxford University Press.