Development of Biomedical Devices for the Extracorporeal Real-Time Monitoring and Perfusion of Transplant Organs

Robert Michael Learney

Department of Bioengineering
Imperial College London
September 2016

A thesis submitted in partial fulfilment of the requirements for the award of *Philosophæ Doctor* (PhD) in Biomedical Engineering and the Diploma of Imperial College London

Imperial College London

Abstract

The goal of this Thesis is to develop a range of technologies that could enable a paradigm shift in organ preservation for renal transplantation, transitioning from static cold storage to warm normothermic blood perfusion. This transition could enable the development of novel pre-implantation therapies, and even serve as the foundation for a global donor pool.

A low-hæmolysis pump was developed, based on a design first proposed by Nikola Tesla in 1913. Simulations demonstrated the theoretical superiority of this design over existing centrifugal pumps for blood recirculation, and provided insights for future avenues of research into this technology.

A miniature, battery-powered, multimodal sensor suite for the in-line monitoring of a blood perfusion circuit was designed and implemented. This was named the 'SmartPipe', and proved capable of simultaneously monitoring temperature, pressure and blood oxygen saturations over the biologically-relevant ranges of each modality.

Finally, the Thesis details the successful implementation and optimisation of a combined microfluidic and microdialysis system for the real-time quantitation of creatinine in blood or urine through amperometric sensing, to act as a live renal function monitor. The range of detection was 4.3μM – 500μM, with the possibility of extending this in both directions. This work also details and explores a novel methodology for functional monitoring in closed-loop systems which avoids the need for sensor calibration, and potentially overcomes the problems of sensor drift and desensitisation.

Dedication

This Thesis is dedicated to my wife, Jennifer Barbara Learney. Your support and kindness over these past 4 years has been unwavering.

With love,

-Rob.

Declaration

I certify that all work presented in this thesis is the product of my own original work unless otherwise referenced.

Dr. Robert Michael Learney

Copyright Notice

Acknowledgements

Martyn Boutelle I would like to sincerely thank you for you support since we met nearly 8 years ago. Without your expertise, experience, and enthusiasm for biosensing, this would not have been possible.

The Dyson Foundation You gave me this exceptional opportunity to explore and develop my own ideas and have some of the most enjoyable times of my life. May you long continue to provide others with the same.

My Colleagues Michelle, Sally, Agnes, Isabelle, Prang, Sahir, Chu, Aidan, Ewan and our newcomer Marsilea, I will remember all of the good times we've shared down in the lab and flying off to conferences. I couldn't imagine a nicer group of people to bounce ideas off over the past 4 years.

My Family You've always been there for me and I thank you from the bottom of my heart.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This Chapter sets the context for this Thesis by introducing the scale and history of organ transplantation.

1.1 The Need for Transplantation

End-stage renal failure (ESRF) is an irreversible and devastating condition for patients and their families alike. Whilst it can arise from a number of causes, the final result is common to all – their kidneys have degenerated to such a degree that they can no longer balance the water and salts required for normal life, nor produce certain key hormones. End-stage renal failure remained a death sentence until the late 1960s, some 20 years after the initial invention of the Kolff hæmodialysis apparatus, because of inadequacies in the technology and facilities of the day.

Sadly, 40 years on from the widespread creation of outpatient dialysis facilities, demand is still growing. This is partly due to natural population growth, but is compounded by the effects of a demographic shift to longer lifespans, and the growing burden of renal injuries caused by diabetes and high blood pressure. Some form of dialysis is therefore a necessary burden for nearly 33,500 patients across the UK.

Approximately one third of all ESRF patient undergo peritoneal dialysis, where a specially prepared fluid is instilled into the abdominal cavity, and drained away again containing waste products normally disposed of via the kidney. This uses the body's own internal membranes to provide the large surface area and selectivity required for dialysis. Many patients will exchange this fluid a number of times each day, and carry on with their normal lives in between exchanges, termed *ambulatory* peritoneal dialysis. Others opt to connect themselves to a bedside device which continuously pumps the exchange liquid into their abdomen throughout the night, so-called *overnight* peritoneal dialysis. A patient is often monitored at a region dialysis centre, or by a specialist nurse, in order to attempt to identify problems with lines or infections before they arise.

The remaining two-thirds of ESRF patients must undergo thrice-weekly hæmodialysis, most often at a specialist centre. This is a time-consuming and costly process wherein patients often need to undergo preparatory surgery to form a permanent access point for direct high-flow blood exchange in their forearm – a *fistula*. Following maturation of the fistula over a number of weeks to months, a patient then visits their local centre, undergoes blood tests, and is connected to a machine for 3-4 hours at a time, before further blood tests and discharge home until the next session.

There has been a recent shift within the NHS to encourage home hæmodialysis over these visits to hæmodialysis centres, largely due to the improvements that patients see in their quality of life, but also because of the cost savings – some £50m per year is spent on just transporting patients to and from their local hæmodialysis centres[1].

The dream for many of hæmodialysis patients is to be able to break free of the cycle of

dialysis and to be able to do something as simple as taking an overseas holiday without weeks of planning for temporary access to a dialysis centre.

Alongside the continuous visits to their dialysis centre, the checkups with their GPs and Renal Physicians, the numerous tablets and injections, the strictly controlled diet and fluid intake, comes the reminder that dialysis slowly kills. Despite all the scientific progress in this field, it is still not possible for most patients with ESRF to achieve a normal life expectancy. In fact the five year survival rate for 18–35 year olds with ESRF on renal replacement therapy is worse than that for some cancers, at less than 60%. These mortalities largely result from cardiac and infection-related deaths.

So not only does ESRF place a severe burden on the patients and their carers, many of whom may find it difficult to contribute to the economy because of their condition and its management, it also places a sizeable financial burden on the health service. In England, the total direct costs of renal replacement therapy (dialysis or transplantation) accounts for some £1.23bn of the total renal care expenditure of £1.64, despite only comprising 2% of the diagnosed renal care population[2].

Alongside these 33,500 patients on dialysis live some 32,700 more patients who have received kidney transplants - the only true 'cure' for ESRF we are currently able to achieve.

Whilst transplantation does bring some costs to patients in terms of dietary restrictions and potential side-effects from immunosuppression, it is incredibly liberating and life-changing. Not only are there benefits for the patients, the health service benefits greatly in terms of annual savings – following the one-off cost of the transplantation surgery, a transplant recipient only costs the NHS around £5,000 per year of transplant function.

It is quite straightforward to see why transplantation longevity is such goal worth achieving – even one extra month of transplant function gives the patient their freedom and saves the health service money.

Unfortunately, the transplant service is in a difficult situation – transplants are still comparatively uncommon, with only 32,700 people in the UK today living with a transplanted kidney. Of the 33,500 patients on some form of dialysis, only 5,275 were on the transplant waiting list as of April 2016, which is a 7% decrease on the previous year, in response to an upwards trend in the number of organ donors over the past 5 years. Some 3,071 kidney transplants were performed in financial year 2015-2016[3], with just over a third of all kidney donations (1035) in the UK being living donor transplants. However, of these, only 83 were donated altruistically to complete strangers. The median waiting time for someone on the list to receive a kidney is 1022 days. This still leaves nearly 259 patients dying on the transplant list each year whilst waiting for a life-saving kidney.

The solution to this problem of supply and demand will probably only fully be met by one

of 3 'cures' – (i) reducing the number of patients with ESRF, (ii) improving the technology for peritoneal or hæmodialysis until it provides a viable life-long portable solution, or (iii) by creating new genetically-matched kidneys in the laboratory, either *de novo* or through depleting the connective-tissue scaffolds of other explanted kidneys and repopulating them with patient-specific stem-cells.

As one can imagine, these possibilities remain quite far from becoming reality.

However, to even reach the state we are in today took many years of research and pioneering experiments, as well as government involvement.

1.2 The History of Renal Transplantation [4–6]

Figure 1.1: 'Saints Cosmas and Damian Healing a Christian with the Leg of a Dead Moor', School of Castille and Leon, ca.1460-1480 - one of many similar representations

Earliest Transplant Experiments

Fantastical accounts of human transplantation date back more than a thousand years, including the (undoubtedly apocryphal) case of the two 3rd century AD Roman Catholic

Saints Damian and Cosmas removing the diseased leg of a white Roman deacon and replacing it with the leg from a black-skinned moor.

In reality, the surgical technology to reliably perform solid-organ transplants only evolved after the development of modern general anæsthesia in the mid-1800s. However, it still took more than 100 years, massive international cooperation and the development of the field of immunology to perform the first successful modern human-to-human transplant.

The earliest cases of attempted renal transplantation serve as interesting historical footnotes of our misunderstanding of biology which many school students could probably now correct. These were heroic mistakes, but ones that needed to be made. Examples include Mathieu Jaboulay's (1860 – 1913) placement of a dog's kidney into the forearm of a dying woman in 1906 in Lyon, and Ernst Unger's (1875 – 1938) use of a monkey's kidney in a dying young girl in 1909 in Berlin. Both of these xenotransplants failed within 24 hours, undoubtedly mediated by hyperacute immunological rejection[1]. This 'biological incompatibility' was recognised by the French surgeon Alexis Carrel (1873 – 1944), who also developed the end-to-end surgical anastomosis techniques used to this day, and for which he won the 1912 Nobel Prize in Medicine.

Following these disheartening attempts, the field stalled for some 20 years until the first true human-to-human transplant by Yurii Voronoy (1895 – 1961) in Soviet Ukraine in 1933. Unfortunately, this was also thwarted by acute rejection leading to the patient dying within 48 hours without producing any urine.

Development of the Field of Immunology

The breakthroughs to answering the cause of Carrel's 'biological incompatibility' began with an unproven hypothesis and skin-grafts performed on the burn victims of World War 2.

In 1949 Sir Frank Macfarlane Burnet (1899 – 1985), a distinguished Australian virologist with a great interest in antibodies, first introduced the immune concept of 'self vs. non-self' and hypothesised that bodies learn to tolerate themselves during embryogenesis by failing to make antibodies against their own cells.

Around the same time, British biologist Sir Peter Medawar (1915 – 1987) and colleagues were attempting to understand skin graft rejection and came across Burnet's hypothesis. They subsequently discovered through that skin-grafts were rejected from incompatible recipients due to a cell-mediated rejection process. In 1953 they also proved that they could induce the immune system of one mouse to 'tolerate' the skin and cells of another

[1]Due to pre-formed circulating antibodies, mostly IgM

by exposing the first mouse to certain immune cells from the second around the time of birth. These experiments and others spawned the field of immunology, and Medawar and Burnet were awarded the Nobel Prize in 1960 for their work.

Resurgence of Interest and Early Immunosuppression

Around 1952, just prior to the discoveries of Medawar and Burnet, an American surgeon by the name of David Hume (1917 – 1973) performed multiple human-to-human kidney transplants with varying levels of success. The longest-lived of his cases was a record 5 months before the young recipient died of accelerated hypertension and heart failure. This and other cases around the same time demonstrated that unmatched transplants were doomed to failure, and that only chance 'partial matching' was the likely reason behind the 5-month longevity of Dr. Hume's success.

Only 2 years after these failures came the first truly successful human-to-human kidney transplant between two 23 year-old identical twins, performed at the same Boston hospital as David Hume by Dr. Joseph Murray (1919 – 2012). Both survived, with the recipient later returning to work, marrying and raising a family. This and 6 similar cases performed over the following years proved the technique was viable, but only between identical twins.

However, this in itself led to a desire to take the findings of Medawar and Burnet and translate them into a way of tolerising potential recipients to imperfectly matched kidneys from unrelated donors.

By 1959, Joseph Murray had extended the possibilities of transplantation to non-identical twins by using pre-operative doses of radiation to cause immunosuppression in the recipient. Two of his patients managed to live normal lives for 20 and 26 years, before succumbing to other illnesses.

Using the same techniques, by 1961 Joseph Murray and colleagues were able to expand their successes to non-twin siblings as well as live unrelated donors.

'Modern' Immunosuppression

The emergence of the current era of immunosuppressive therapy began in London in 1959 with work by British surgeon Sir Roy Calne (1930 –). Recognising that the earliest stages of an immune response against 'foreign' tissue involves a rapid proliferation of T and B-lymphocytes, work began using 'anti-metabolites' – drugs that would inhibit DNA synthesis and thus halt cell proliferation around the time of transplantation. It was hoped

that this would also lead to tolerisation of the new organ by the recipient immune system, a hope that has unfortunately not been entirely borne out.

Using a then-experimental drug 6-mercaptopurine, Calne demonstrated prolonged kidney transplant survival in dogs. The drug was then used to complement total body irradiation, but unfortunately proved to have a very narrow therapeutic index[1].

Calne then visited the Boston hospital where Joseph Murray worked in order to continue experimenting with a less toxic pro-drug form of 6-mercaptopurine, azathioprine. This is metabolised within working cells into 6-mercaptopurine, delivering the drug directly to the nucleus. After a successful series of dog experiments, Calne returned to London in 1961 and combined azathioprine with steroid treatments (pioneered just one year prior by Willard Goodwin (1915 – 1998) at UCLA) to create a peri-transplant pharmacological immunosuppression regimen which remained in use for nearly 20 years. This pharmacological breakthrough allowed surgeons to abandon total body irradiation and to push the boundaries of kidney transplantation even further.

Calne later pioneered the use of another immunosuppressive agent, ciclosporin, in the 1970s. This agent also works by preventing T cell activity, but through a more selective mechanism, acting only on those which have been stimulated to respond to foreign tissue.

The First Cadaveric Renal Transplant

April 1962 could be considered as the ‘dawn’ of renal transplantation as performed today, when Joseph Murray performed the first successful transplantation of a kidney from an unrelated deceased donor using pharmacological immunosuppression. The patient lived for over a year.

After a flurry of activity based on this news, a small international conference was convened in Washington DC in 1963. Unfortunately, the results showed that whilst the donor pool was now greatly expanded, only 19% of cadaveric recipients had survived, compared with 48% using the more limited resource of related donors. This led Joseph Murray to conclude “Kidney transplantation is still highly experimental and not yet a therapeutic procedure."

The reasons for these poor survival rates are interesting and open up a new side to the history of renal transplantation.

[1] *Therapeutic index* – the ratio between the useful and toxic or lethal dose

Availability of Organs

With proof that it was now possible to use unrelated deceased donor kidneys to greatly change the lives of individuals suffering from end-stage renal failure, the earliest pioneers were unfortunately frustrated by the lack of recognition and definition of 'brain death'. Until a definition and appropriate tests emerged in the mid 1960s, death was taken to be the point at which the heart stopped beating.

This meant that in reality, throughout the 1950s and early 1960s, once a patient had entered a state of *coma depassé)* (beyond coma, or 'irreversible coma') and mechanical ventilation was discontinued, it could take hours from the time the patient's heart stopped beating for them to be declared legally deceased, wheeled into the operating theatre, prepared for surgery, to have their kidney exposed, clamped and extracted and finally have it prepared for re-implantation in a desperately ill recipient. All this time the kidney would be at body temperature, yet starved of circulating blood with its vital oxygen and nutrients.

In 1963, Belgian surgeon named Guy Alexandre performed the first ever transplantation using the kidney of a patient in 'irreversible coma' without stopping mechanical ventilation. This meant the kidney remained supplied with blood from a beating heart for all but the shortest time between clamping, extraction and reimplantation. By 1966, Dr. Alexandre had used 9 such patients as donors for renal transplantation.

Announcing these results and the appropriate tests to define 'brain death' to a 1966 symposium in London, the audience in general was highly suspicious of the news. Sir Roy Calne said "Although Dr. Alexandre's criteria are medically persuasive, according to traditional definitions of death, he is in fact removing kidneys from live donors. I feel that if a patient has a heart beat, he cannot be regarded as a cadaver."[7]

Not two years later, the Ad Hoc Committee of the Harvard Medical School released their report on the 'Definition of Brain Death', largely agreeing with Dr. Alexandre's opinion and criteria[8]. Over the following few years, laws were changed in Europe and the USA to reflect the new medical thinking, finally permitting widespread use of heart-beating but brain dead donors. There remains a subtle difference between the UK and the USA in the diagnosis of brain death, with the UK requiring only brain-stem death, and the US requiring global or *whole brain* death.

Differences in Donor Organs

Not every organ donor is created equal, and the source of an organ has a profound impact on its suitability for transplantation. As discussed in Section 1.1 on page 23, the two major sources of organs in the UK today are from living and deceased donors, with two further

subdivisions within each of these, namely live related donors and live unrelated donors, and donors who are Deceased after Brain-stem Death (DBD) or Deceased after Cardiac Death (DCD). These divisions are illustrated in Figure 1.2 below:

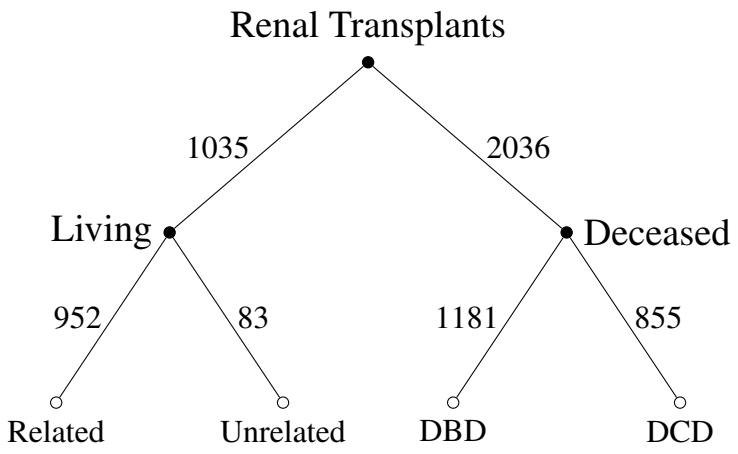


Figure 1.2: Relative proportions of the sources of donor kidneys 2015-2016[3]

By and large, well-matched organs from living donors provide the best outcomes, with unadjusted figures showing 94% of recipients surviving for 5 years compared against 87% for recipients of kidneys from deceased donors[9].

However, the majority of recipients must await the death of another person before they can receive an organ. In the UK, these most commonly result from DBD rather than DCD donation. Some countries, such as New Zealand, Germany or Ireland, have no means to support DCD organ retrieval and must rely solely on DBD organs.

Deceased after Brain-stem Death (DBD)

In the UK, these organs are retrieved from individuals who have suffered catastrophic and irreversible damage to their brain stem, leading to the loss of function necessary to support life.

Brain-stem death is diagnosed in a hospital through neurological testing. The donor is usually maintained on a respirator in an intensive care unit with no chance of return to consciousness, but with the heart beating spontaneously and the other organs functioning normally. There may or may not be a requirement to support the donor's circulation with drugs and fluids. For this reason, DBD donors are sometimes referred to as *beating-heart cadavers*.

As a result, the kidneys from DBD donors are usually functioning normally at the point the donor is taken to the operating theatre, and throughout the procedure during which their

organs are retrieved, cooled and transported to various receiving units around the country, where recipients are waiting. Note that the adrenal glands are separated from the kidneys at the time of retrieval, and are discarded.

Because these organs are oxygenated and perfused for longer within the donor, and the retrieval occurs in a predictable and controlled fashion with a minimal warm ischæmic time, it is no surprise that these organs tend to function earlier and survive for longer in well-matched recipients than those retrieved after uncontrolled cardiac death[10].

These donations can be planned in advance, including virological screening and tissue typing, leading to a minimal period of cold ischæmia en route between donor and recipient sites. DBD organs are also offered nationally, ensuring that the best possible matching recipient can be found for an optimal long-term outcome.

Deceased after Cardiac Death (DCD)

There are two types of DCD donation – controlled, and uncontrolled. These are further divided according to the modified Maastricht criteria which describe the circumstances and the location of death[11, 12].

As can be imagined, uncontrolled DCD means a person has died unpredictably, perhaps at home, in the ambulance whilst being transported to hospital, or on a ward (including intensive care). These patients may have multiple comorbidities which resulted in their sudden and uncontrolled cardiac death. Those who arrive in the emergency department will usually be recieving cardiopulmonary resuscitation with no spontaneous cardiac or respiratory function. Standard protocol calls for CPR to be continued until the family consents to donation, and a retrieval team arrives. At this point, resuscitation is stopped and the team waits for a period of up to 10 minutes to satisfy the definition of death and to allow the family a moment to grieve. After this period, organ retrieval proceeds in a rapid fashion in the operating theatre.

Most DCD donations within the UK are instead controlled, with donation planned in advance. In this cohort, the patient is usually in a hospital intensive care unit following a catastrophic brain injury where survival is unlikely. After consultation with the family, respiratory support is withdrawn, leading to apnoea and hypoxic cardiac arrest over a number of minutes to hours. This is accompanied by a waiting period to satisfy the definition of death, after which the family is allowed to grieve. Retrieval then proceeds in the operating theatre. Similar to planned DBD retrieval, consent can be obtained for virological studies and tissue typing in advance to reduce the cold ischæmic time during identification of a potential recipient for advanced tissue typing[13].

Unfortunately, whilst DCD donors outnumber DBD donors, only 64% of them proceed

to organ retrieval, compared against 97% suitability for retrieval in DBD donors. This is largely a function of the prolonged time to death after withdrawal of treatment. Furthermore, only 85% of kidneys retrieved from DCD donors are deemed suitable for transplantation, compared to 92% of those from DBD donors[3], largely owing to donor factors including infections, cancers, arterial disease, organ anatomy and the donor's pre-morbid condition, including age and obesity.

DCD organs are also only offered on a local or regional basis in the UK, due primarily to the lack of established national allocation protocols. This means that there is a smaller pool of potential recipients with a subsequently higher theoretical risk of a poorer match when compared against the larger national recipient pool.

Kidneys retrieved from either type of DCD donor also have a higher risk of *primary non-function* and *delayed graft function* versus DBD kidneys (51% vs. 24%) owing to the prolonged period of warm ischæmia prior to retrieval, and the peri-mortem combination of hypotension, hypoperfusion and hypoxia in this mechanism of donation. However, DCD kidneys which recover from this initial period of insult do appear to result in long-term organ survival rates comparable to those of DBD kidneys[3, 10].

Advances Since 1970

By the end of the 1960s, the groundwork for renal transplantation was firmly established and the concepts were accepted across much of the Western world. Surgeons were able to give patients a predictable and reliable quality of life following transplantation, with access to organs from a wide pool of donors.

Our understanding of immunology improved further with a landmark 1968 publication in the New England Journal of Medicine, demonstrating the benefits of matching donors to patients according to serotype using an early form of the HLA (human lymphocyte antigen) classification scheme[14].

Throughout the 80s and 90s there emerged multiple new pharmacological agents to modulate the immune response of the organ recipient in order to prevent and reverse not only acute but also chronic rejection. It is now a clinical fact that 90% of patients can expect to live at least 5 years following a DBD kidney transplant, and 88% following a DCD transplant[3].

Finally, and perhaps one of the greatest contributors to the continuing success of organ transplantation, the NHS Organ Donor Register was launched on the 6th of October 1994. As of 2016 there are nearly 21 million people on the Register, representing nearly a third of the UK adult population.

Other countries have gone further and introduced compulsory registration, with an active 'opt-out' or 'presumed consent' register, rather than our 'opt-in' method. This remains a controversial subject in the UK, and the most recent review of the law in 2008 recommended that the current system be retained for the time being.

The most recent development within the UK was the first altruistic donation of a kidney from a living donor to an unrelated person in 2007. This source of organs is continuing to grow over time.

1.3 History of Extra-Corporeal Preservation

The problem of preserving a donor kidney in the best possible physiological condition has concerned transplant surgeons since the field first expanded in the early 1960s.

However, the idea of preserving an organ outside its natural body is not new to medicine. Some of the earliest experiments with organ preparations date back to Carl Friedrich Wilhelm Ludwig's (1816 – 1895) perfused frog hearts in the 1860s[1]. Building on this early work, Sydney Ringer (1835 – 1910) discovered in 1883 that specific concentrations of calcium, potassium and sodium ions were necessary for the prolonged function of isolated organs[2].

These early preparations lay the groundwork for the likes of Alexis Carrel and Charles Lindbergh (1902 – 1974), who in 1937 established that it was possible to use artificial pumps and heat exchange loops to mimic normal physiology and support entire isolated organs (in their case, the thyroid glands of cats) for 3 – 21 days[15].

In a quite unrelated early series of experiments to study the means by which the kidneys produce urine from blood plasma, Reginald G. Bickford (1913 – 1998) and Frank Winton (1894 – 1985) serendipitously discovered in 1937 that isolated dog kidneys survive many times longer when cooled below a critical temperature, and are more resistant to metabolic insults such as cyanide poisoning and reversal – an early 'ischæmia-reperfusion' experiment[16].

This concept was apparently overlooked during the early stages of the development of transplantation, probably because clamping time was minimal, and the only transit of the organ was between neighbouring operating theatres. Nevertheless, it was revisited in 1960

[1] Ludwig was one of the earliest 'great pioneers' of modern physiology, inventing various apparatuses for studying gas exchange in muscles, stimulation of nerves and recording the blood pressure within living vessels, whilst also caring for the pain and suffering of his experimental animals. His rejection of the earlier ideas that special and indecipherable laws governed biology could well be said to have opened up the entire field of modern medicine.

[2] Apparently by accident when his laboratory technician swapped his normal distilled water for simple London tap water

by Lapchinsky, a Soviet surgeon interested in studying how best to preserve organs and limbs for transplantation. Lapchinsky was able to report the successful storage of canine kidneys at 2°C – 4°C for up to 28 hours using refrigerated oxygenated blood.

In 1963, the pioneering transplant surgeon Sir Roy Calne famously demonstrated that it was possible to excise kidneys from dogs, preserve them in chipped ice for 7 – 12 hours and subsequently re-implant them with good return of function[17], a protocol nearly identical to the current method of cadaveric retrieval. This paper also covers some of the key evidence known at the time which remains true to this day – that renal oxygen consumption falls to less than 5% of normal at 5°C, as does the active tubular transport of water and salts.

Thus began the era of static cold storage (SCS) for extracorporeal renal preservation. By suppressing renal metabolism, it was now possible to maintain kidneys outside the body for 12 – 24 hours with good return of function once re-implanted, long enough for centre-to-centre transportation. Certain aspects of the protocol have been refined over time, such as flushing the freshly-explanted kidney with cold preservation solutions, or heparinised autologous blood prior to packing in sterile ice.

1.3.1 Drawbacks of Static Cold Storage

Figure 1.3: A 'Classic' Static Cold Storage Box

Whilst SCS dramatically improved the availability and physiological condition of transplant organs, it is not without certain severe drawbacks – not least of which is that kidneys maintained in this manner do continue to deteriorate over time. SCS remains in use for only short (around 24 hours) periods of preservation, with a dramatic fall in function at even 72 hours. Compare this to the 80+ years over which a normal human body is able to maintain good renal function.

Cooling to low temperatures also causes harmful changes to cellular biochemistry. Firstly, whilst oxygen consumption does fall to ≈5% of normal around 5°C, it does not fall to 0%. There remains a basal oxygen requirement which must therefore be satisfied if cells are not to die. Unfortunately, in a static state the only source of oxygen is diffusion from any residual flushing solution or from the atmosphere through the parenchyma.

It is possible to roughly calculate the oxygen deficit in such SCS kidneys. Given that renal oxygen demand in physiologically normal human kidneys equals approximately 10mL O_2/minute (measured *in vivo* via intravascular catheterisation techniques[18, 19]), at 5% of normal metabolic rate, SCS kidneys therefore require ≈0.5mL/minute.

Pure water at 5°C has a maximum O_2 solubility of 12.77mg/L, which only decreases in the presence of the salts necessary to balance the tonicity of the solution. Assuming a 50mL flush through the vasculature of the organ prior to cold packing, this makes available only a total of 0.63mg or ≈0.44mL O_2 via the vasculature. The SCS kidney must therefore rely on atmospheric oxygen diffusing through the parenchyma for its basal metabolic requirements – a very inefficient process, compounded by the decreasing diffusivity of oxygen through water at lower temperatures, clarifying why the survival of these kidneys beyond even 12 hours is strictly limited.

However, the metabolic disruption of cryostorage is even more pronounced when one considers the suppression of the membrane-bound Na^+ pump at low temperatures, but the continued functioning of the Na^+/K^+–ATPase[20]. This leads to the consumption of ATP without mitochondrial oxidative phosphorylation able to restore the resting levels, an ionic and osmotic imbalance across the cell membrane leading to the loss of K^+ and Mg^{2+} ions, and therefore Na^+ and H_2O accumulation with concomitant cellular œdema.

Lowering the temperature of the cell membrane also alters the fluidity of membrane micro-domains, much like placing olive oil into a refrigerator. This affects the functioning of their resident enzymes, whether mitochondrial, nuclear, or as part of the endoplasmic reticulum. These changes affect protein synthesis, free-radical scavenging, antioxidant regeneration and ATP generation. An excellent summary can be found in the review by Taylor *et al.* (2010)[21].

1.3.2 Development of Cold Storage Solutions

Research efforts to counter these effects in SCS have largely involved the creation of storage and flushing solutions with specific compositions. One common philosophy in this area is to bathe the kidney in a solution which mimics the intracellular K^+ concentration, thereby neutralising the activity of the Na^+/K^+–ATPase and reducing ATP consumption and thus O_2 demand further than by hypothermia alone. Other chemicals are added to the

solution, including antibiotics, neutral osmotic agents such as raffinose to increase tonicity and resist cellular oedema, antioxidants such as lactiobionate for scavenging free radicals, and occasionally ATP precursors for non-oxidative ATP regeneration.

The most successful of these is the *University of Wisconsin Solution* (UW), which has been adopted as the 'gold standard' for preservation since the late 1980s, and has proven superiority over other 'extracellular'–type solutions for all solid organs[22]. However, at a cost to the NHS of £116/litre, and the requirement for multiple litres per kidney, the NHS uses a slightly different solution known as 'Marshall's hypertonic citrate' for approximately ¾ of renal transplants, which only costs £9.60/litre[1].

One interesting modification to the composition of the UW solution in the past decade has been to exchange its hydroxyethyl starch (HES) for a polyethylene glycol (PEG). HES solutions have been shown to cause red cell aggregation, compared to the morphology-preserving characteristics of high molecular weight PEGs[23]. This has been reflected in the improved early functioning of kidneys preserved with this new *IGL-1 Solution* (named for the *Institut Georges Lopez*) in France[24].

Yet even with these exotic solutions, preservation is still rarely able to exceed 48 hours with good, reliable outcomes.

1.4 History of Machine Preservation

Briefly introduced on page 32, the earliest experiments with maintaining solid organs outside the body were often carried out at normal body temperature, using recirculating autologous blood. However, once clinical transplantation took off in the 1960s, with the known benefits of cryopreservation on ischæmic organs, existing pump and oxygenation technologies were soon adapted and applied to maintaining whole organs at lower temperatures[25].

The period from 1964–1968 saw a number of publications emerge from Dr. Alistair L. Humphries Jr. at the Medical College of Georgia in Augusta, demonstrating that it was possible to maintain explanted canine kidneys for up to 5 days (according to one publication) with good functional outcomes, using a device to continuously recirculate diluted plasma.

However, the greatest breakthroughs in the hypothermic machine preservation (HMP) of human kidneys came from Dr. Folkert O. Belzer (1930 – 1995), working out of the University of Wisconsin. His work around the same time of Dr. Humphries led to the creation of the first portable, self-contained, HMP device.

[1]NICE technology appraisal guidance 165, issued January 2009

Figure 1.4: Belzer's Perfusion Machine, from Fuller *et al.* 2007[25]

The large size of early devices in the pre-microchip era was dictated by that of the best available equipment of the day. Nevertheless, the device proved sufficiently portable that in a 1971 episode of showmanship and medical necessity, Belzer was able to transport a kidney from a donor in San Francisco to the Netherlands, and into a waiting recipient some 37 hours after retrieval.

The LifePort® Kidney Transporter[1] is the latest HMP device to be brought to market, and has the best supporting evidence to date (see page 38 for a discussion). The latest version of this device incorporates a specifically-designed kidney retrieval tray and cannula system for rapid attachment to the machine's recirculation system, a miniature roller pump and bubble trap, a well as a heat exchanger, GPS tracking system and live digital readouts of certain parameters (flow rate, perfusion pressure and temperature, amongst others).

[1]Organ Recovery Systems, Des Plaines, IL, USA

Figure 1.5: The LifePort® Kidney Transporter from Organ Recovery Systems

1.4.1 Issues in Hypothermic Machine Preservation

The development of HMP has not been entirely free of problems and controversies. A number of issues have been explored by multiple groups worldwide in the 40 years since Belzer's device was first demonstrated.

It became clear that the earlier use of whole or diluted blood was not viable, partially due to hæmolysis in the pump and oxygenator devices of the day, or due to the formation of micro-thrombi which could occlude the renal vasculature. Furthermore, shearing forces from tubing, or frothing from bubble oxygenation appears to denature plasma proteins and lead to a variety of negative side-effects including damage to the vascular endothelium. In fact, shear forces applied to vascular endothelium under low-temperature conditions may be harmful in and of itself[25], perhaps for reasons of altered membrane fluidity (see page 34 for details).

Developing an appropriate perfusion solution was therefore paramount to the success of the technology. Many of the same principles as discussed in Section 1.3.2 on page 34 apply to perfusion as well as static systems. The earliest work with diluted plasma also led to the discovery of soluble lipoproteins which precipitate out of solution during rapid thawing from frozen. Removing these greatly reduced the damage to perfused kidneys.

Later, Belzer would go on to co-develop the fully-synthetic UW solution in 1987, which greatly changed the success of both SCS and HMP for kidneys. Research continues to this day on optimising synthetic perfusion solutions, particularly the osmotically active component of the solution using compounds such as mannitol and polyethylene glycols in the place of UW's hydroxyethyl starch.

Interestingly, no existing commercial solutions contain oxygen carriers for HMP. There have been some experiments performed using perfluorocarbon emulsions[26], but with

the ability of even a small amount of residual perfluorocarbon to stimulate the reticulo-endothelial system, their resistance to enzymatic digestion and therefore necessity for elimination through the lungs have resisted their uptake in practice, and slowed or stalled their progress through standardised clinical trials.

Therefore, all existing HMP solutions rely on the O_2 solubility of water to meet the needs of the target organ. The rate of delivery of solution to the organ is therefore critical to meet its oxygen demand and flush out toxic metabolites.

HMP pumps were initially developed to generate oscillatory, pulsatile perfusion waveforms in order to mimic natural cardiac output. However, it has also been found that maintaining a single mean pressure is just as successful. Perfusion pressures for pulsatile HMP systems for kidneys have been widely established at $^{60}/_{40}$mmHg[25], equivalent to a mean perfusion pressure of ≈47mmHg. Any higher than this seems to risk interstitial oedema.

Finally, early plastics used in the connection tubing also contained large amounts of phthalate plasticisers, which at even low levels can be toxic to mitochondria[20], and may have even been responsible for a large number of historically poor outcomes with HMP.

1.4.2 Current Use of Hypothermic Machine Preservation

Unfortunately, despite the promise of HMP to maintain kidneys in better condition for longer than SCS, this area remains controversial.

The most convincing evidence to date comes from the 3-year follow-up report to an international, randomised controlled trial commenced in 2008, directly comparing cold storage to machine perfusion[27]. The study's design was extremely robust, with more than 300 paired kidneys retrieved from deceased donors (DCD and DBD) randomly split with one placed into cold storage and the other machine perfusion. There was a clear and statistically significant advantage to graft survival from machine perfusion (91% surviving at 3 years, vs. 87% for SCS) overall, and DBD kidneys specifically (91% vs. 86%), but not DCD kidneys.

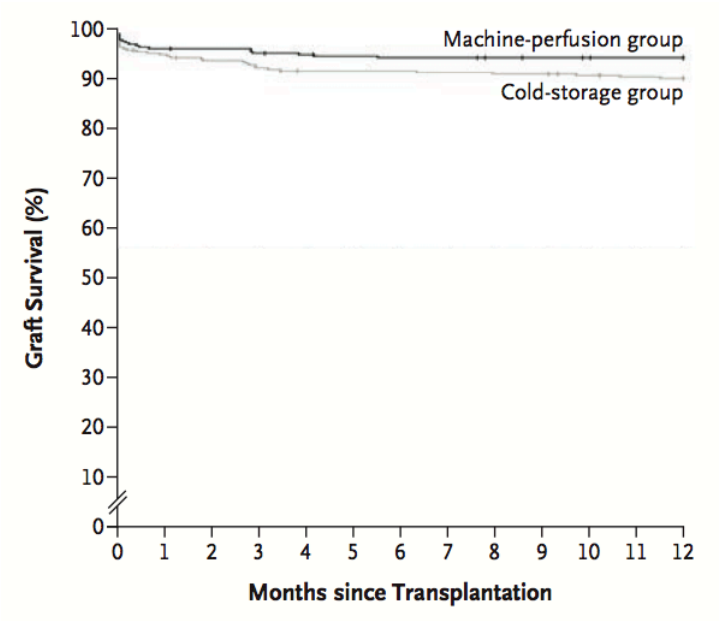


Figure 1.6: Graft Survival: Machine Perfusion vs. Static Cold Storage. From Moers *et al.* 2012[27]

Each percentage improvement equates to surviving kidneys, improved lives and money saved by healthcare services.

However, clinical practice in the UK has yet to take full advantage of HMP, perhaps due to perceived cost issues. The British Transplantation Society believes that only 2% of kidneys from deceased donors were stored using machine perfusion between 2000 and 2007. However, a 2009 study by NICE remained neutral on advising between SCS or HMP (using the LifePort® system)[1]. This advice has not yet been revised.

1.4.3 A Paradigm Shift

With great strides having being made in automated control systems, precision manufacturing and polymer technologies since Carrel and Lindbergh's work in 1935, it seems unusual that the idea of warm machine perfusion had been almost totally abandoned since cold perfusion technologies appeared in the 1960s.

Hypothermic preservation, both SCS and HMP, seem to have become popular simply because the underlying philosophy was straightforward – 'cool the kidney to reduce metabolic demand and thereby resist ischæmic injury' – and adequate results were achievable with minimal technology. Research into cryobiology, preservation solutions and miniaturisation of machine perfusion systems seem to have developed over the intervening years in order to complement the original philosophy rather than to challenge it.

[1]NICE technology appraisal guidance 165, issued January 2009

It is perhaps time to revisit some of the older ideas. After all, we know a warm, well-functioning and immune-matched human body is the best place to preserve and resuscitate kidneys.

Furthermore, unlike SCS or HMP, warm perfused kidneys create urine and behave like normal kidneys in-vivo, carrying out filtration and ion balancing functions, and presumably hormonal functions in the longer term. This would provide a perfect environment for expanding a whole range of possibilities which are presently constrained by the need to implant and perfuse an organ as quickly as possible. These include:

- Testing a donor organ prior to implantation to gain some sense of its likely function in the recipient
- Reconditioning an injured kidney to restore its normal function
- Buying time to immunotolerise the organ or the recipient prior to implantation

The number of possible therapies or studies that could be enabled by a longer duration of extracorporeal preservation without compromising function, and perhaps improving it, could be extensive. A successful testing and rejuvenation strategy could also greatly expand the pool of potential DCD donor organs.

Is it not therefore possible and desirable to employ modern pumping technologies, dialysis membranes and membrane oxygenators alongside intelligent biosensing and control systems in order to at least partially replicate human physiology in a small portable system?

1.4.4 Warm Machine Perfusion

There has been a resurgence of interest in warm machine perfusion (WMP), or normothermic perfusion (NP) since the mid-1990s. Two of the most active groups in this area are led by Dr. Lauren Brasile, the Director of Research Laboratories for the Department of Surgery at Albany Medical Centre, New York, USA [1] and Dr. Sarah A. Hosgood of the University of Cambridge, UK.

More than 15 years of work by the Brasile group has demonstrated that it is possible to preserve cadaveric human kidneys at 32°C for up to the same length of time as existing SCS and HMP solutions[28]. Moreover, extracorporeal warming of SCS kidneys prior to re-implantation appears to ameliorate the function of ischæmically injured kidneys[29, 30], a feat which is not possible with hypothermic machine perfusion.

Later work by the same group demonstrated the ability of using WMP to repair

[1] Dr. Brasile also heads a company called Breonics, Inc. (Schenectady, New York, USA) in order to commercialise her discoveries

ischæmically injured or other marginal kidneys using a perfusion medium containing growth factors (fibroblast growth factor)[31], or to 'immunocloak' the kidney in order to reduce immunogenicity after reimplantation (using a 'nano-barrier membrane' to coat the vascular endothelium)[32]. Most of the reimplantation work has been performed using canine kidneys, with human kidneys only being subject to histological examination and not reimplantation.

Despite publishing widely, Brasile's commercial entity Breonics, Inc. has yet to release a device to the market which uses their patented 'EMS' (Exsanguinous Metabolic Support) technology. EMS is a cell-free perfusate based on an enriched tissue-culture medium containing 6% w/v polymerised bovine hæmoglobin, which is circulated through an oxygenation circuit using a pulsatile pump, and monitored with O_2, CO_2 and pH sensors. Much like HMP, the EMS perfuses the kidney at a lower than normal temperature and flow rate, and at a pressure of only $^{50}/_{30}$ mmHg[29].

Unlike the Brasile group, the work by Hosgood *et al.* uses a commercially-available pædiatric cardiopulmonary bypass machine (Medtronic, Watford, UK). In a 2011 study using oxygenated autologous blood passed through a leukocyte depletion filter to reduce immunogenicity and recirculation-induced activation, porcine kidneys were perfused for up to 22 hours at either true normothermia (37°C) with a mean arterial pressure of 75mmHg, or perfused with UW solution at 2°C – 4°C and 30mmHg on a LifePort® Kidney Transporter prior to auto-reimplantation. The equal survival of both groups demonstrated the possibility of using unmodified whole blood and true normothermia for renal preservation[33].

The Hosgood group also broke new ground in 2011 when they were the first group in the world to successfully transplant a human kidney resuscitated by near-normothermic perfusion (at 34°C) prior to transplantation. The study is made even more important by the use of the contralateral donor kidney as a control, subjecting it only to SCS. Both kidneys were preserved by SCS for at least 10 hours, after which one kidney was randomly selected to undergo WMP for 30 minutes prior to transplantation. The second kidney was maintained with SCS for a further 3 hours during delivery to its patient.

The dramatic effect of this brief period of oxygenated, normothermic perfusion can be seen in the figure below, taken from the paper[34]:

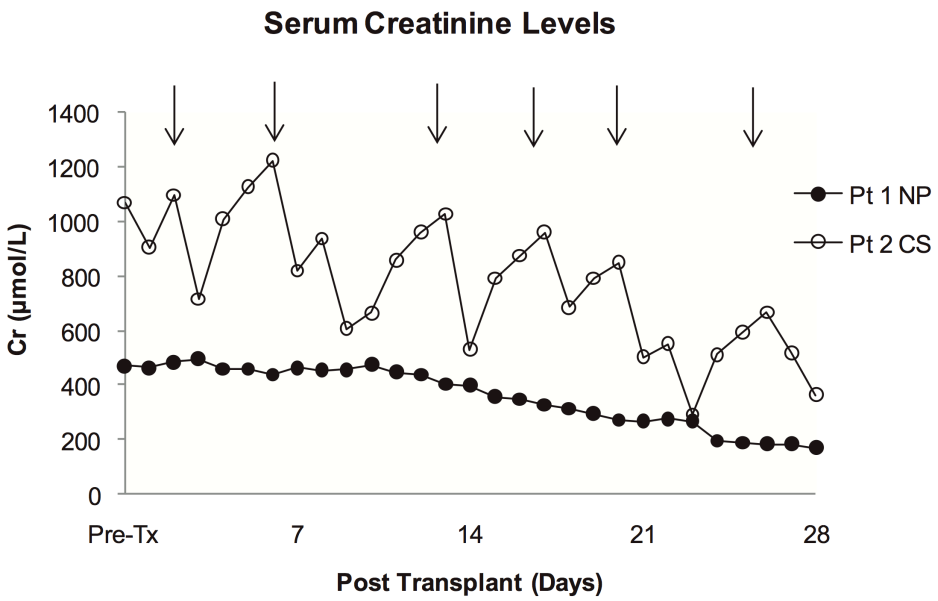


Figure 1.7: Serum Creatinine Levels Following Transplantation. NP – patient's kidney underwent normothermic perfusion. CS – patient's kidney maintained by static cold storage. Lower results are better. Arrows indicate need for dialysis therapy in the CS patient.

These results were followed by the announcement in March 2013 of another first-in-man transplantation following warm perfusion, this time of a liver, maintained by a pre-commercial device invented at the University of Oxford.

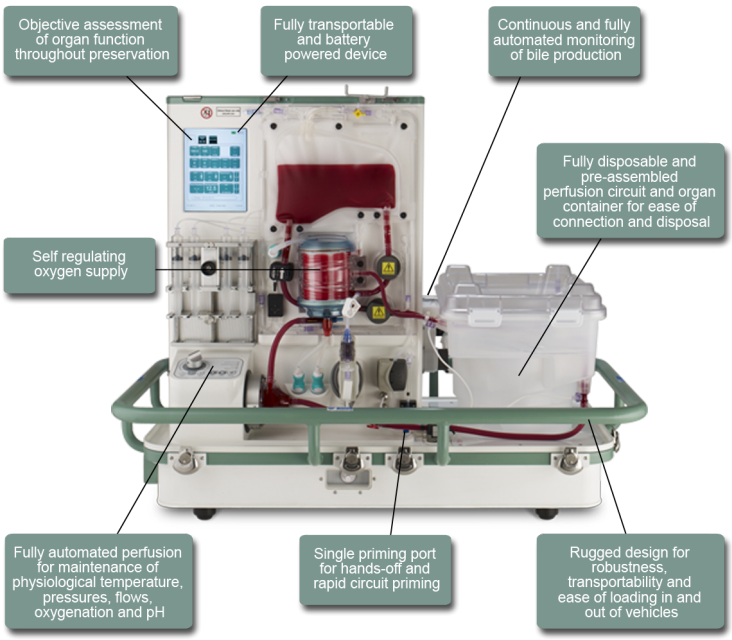


Figure 1.8: The OrganOx® *metra*™, OrganOx Ltd, Oxford, UK

The Hosgood group has since successfully resuscitated two kidneys initially deemed unsuitable for transplantation with an hour of normothermic extracorporeal perfusion,

achieving good results post-implantation[35]. In this particular study, they also claim that up to 19 rejected DCD kidneys from a sample pool of 22 may have been suitable for transplantation, which would greatly the expand the potential donor pool from this population.

1.4.5 Project Proposal

It therefore appears to be a highly appropriate and exciting time to study warm machine perfusion for extracorporeal organ preservation.

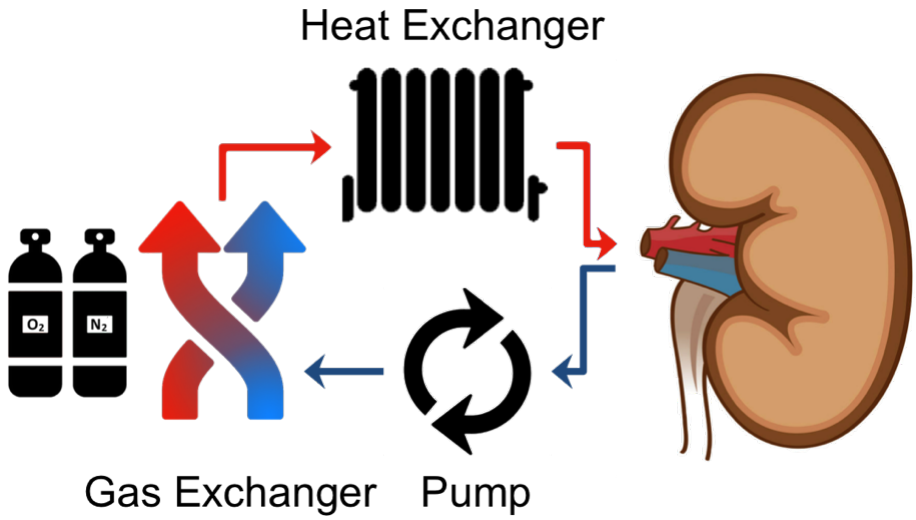


Figure 1.9: The basic arrangement of a normothermic blood perfusion circuit showing the pump, oxygenator and heat exchanger

A perfusion circuit designed to support a kidney by mimicking the *milieu interieur* of the human body must be able to reproduce the basic homeostatic functions as closely as possible, including:

i The delivery of adequate oxygen and nutrients to the organ
ii Clearance of waste products
iii Ability to maintain the target organ at the normal internal temperature of the human body

Fortunately a normally functioning kidney is largely able to clear its own waste products, which only requires the system to provide a means of nutrient delivery, gas exchange and circulation.

The main thrust of my PhD work will be to use design principles and the best current technology in order to engineer the components required for a self-contained, portable device that is able to maintain a kidney outside the body whilst monitoring a wide range of physiological parameters. Other goals will be to minimise the cost of any final devices, and to keep them as simple as possible for non-experts to use for the first time.

A number of benefits could arise from a device that is able to maintain function for even slightly longer than the maximum 48 hours achievable with current technology. The earliest pioneers in transplantation dreamed of 'organ banks'[25], much like the blood banks in our

hospitals today. Belzer also showed us that an international exchange programme could be possible if we are able to keep organs alive for long enough.

Longer periods of warm extracorporeal perfusion with normal kidney function could allow for better preparation of both the organ and the recipient, as well as more robust matching protocols or even full tolerisation therapy for the recipient to avoid lifelong immunosuppression.

Work by Brasile and Hosgood *et al.* has already shown that it is possible to allow kidneys to recover and self-repair after a period of ischæmia, and to provide a window during which other interventions can be applied, such as stem-cell therapies[36], immunocloaking[32], or adhering antithrombotic proteins directly to the endothelium[37].

Finally, if nothing else, a small, portable, whole-organ support system would be very useful to pharmacological research and the emerging science of growing or 3D-printing organs in a laboratory.

Chapter 2

Normal Renal Physiology

This Chapter briefly covers the topic of normal renal physiology, which provides the basis for many of the design decisions discussed in later Chapters.

2.1 Introduction to Renal Physiology

In order to properly assess, imagine and improve methods of warm machine perfusion for isolated kidneys based on our modern scientific understanding of renal physiology, we will begin with a brief review of the state of this knowledge.

Following this, we will discuss the homeostatic mechanisms by which the kidney itself is maintained within a healthy human body.

After all, a well-functioning human body can maintain the kidney for the entirety of a person's life, some 70–80 years in the developed world. Yet even with the best tools and clinical knowledge that we have to hand, we can barely achieve a few days' preservation of function in a kidney maintained outside the body.

Our current understanding of the function of the kidneys has developed through a long process of anatomical and physiological study throughout the 19th & 20th centuries. We now know that the kidneys perform 4 major homeostatic functions[38]:

1. Regulation of the circulating blood volume
2. Regulation of the ionic composition of the extracellular space
3. Excretion of water-soluble metabolic wastes
4. Synthesis of hormones responsible for blood pressure, oxygen transport and Ca^{2+} metabolism

2.2 The Formation of Urine

This section will begin with a review of the mechanisms by which the kidney regulates the composition and volume of blood in circulation and the excretion of wastes via the urine. Note that only kidneys undergoing warm blood perfusion will generate urine. Those maintained by static cold storage or undergoing cold machine perfusion do not. This creates additional complexity for any closed-loop warm perfusion system, as any circulating volume lost as urine will have to be replaced by an appropriately-designed isotonic salt solution.

The Nephron

The nephron is the minimal functional unit of the kidney. There are between 800,000 and 1,500,000 individual nephrons in each kidney. The loss of individual nephrons occurs naturally over an individual's life, but it is the accelerated and/or disproportionate loss of

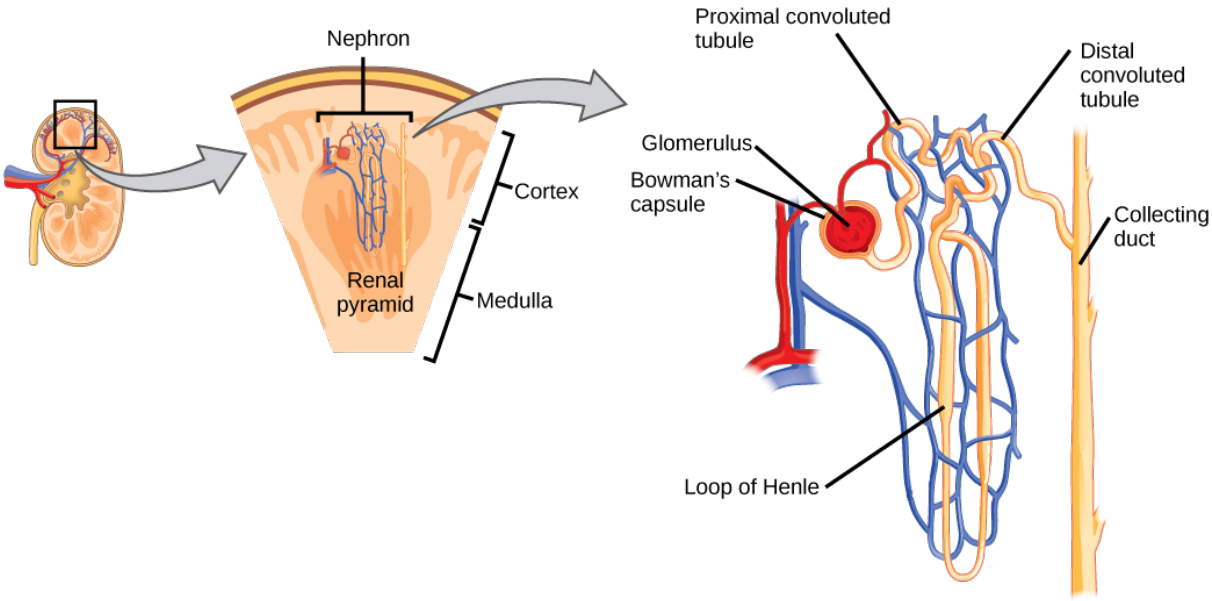


Figure 2.1: The nephron as the functional unit of the kidney. OpenStax [39]

nephrons that leads to renal disease. The causes of such accelerated loss are multifactorial and too extensive to discuss in this Thesis.

The rest of this section will discuss the role of each part of the nephron in the formation of urine, beginning with initial filtration at the *glomerulus*.

The Glomerulus

Urine begins as an ultrafiltrate of blood, with the exclusion of the majority of circulating components such as erythrocytes, albumin, immunoglobulins and other high molecular weight plasma proteins (>35kDa) by the negatively charged glycoproteins of the basement membrane between the interdigitating podocytes of Bowman's capsule. The capillary loops within these capsules are termed the *glomeruli*, from which we derive the term that presently defines overall renal function – the 'Glomerular Filtration Rate' (GFR) of the blood – in ml/minute.

The concept of renal filtration as a marker of renal function is covered later in Section 2.7.1 on page 68.

Lying beside the glomerulus are *juxtaglomerular cells* and the cells of the *macula densa*. Together these comprise the *juxtaglomerular apparatus* (JGA) which plays a role in the hormonal function of the kidney for modulating blood pressure. This is covered further on page 54.

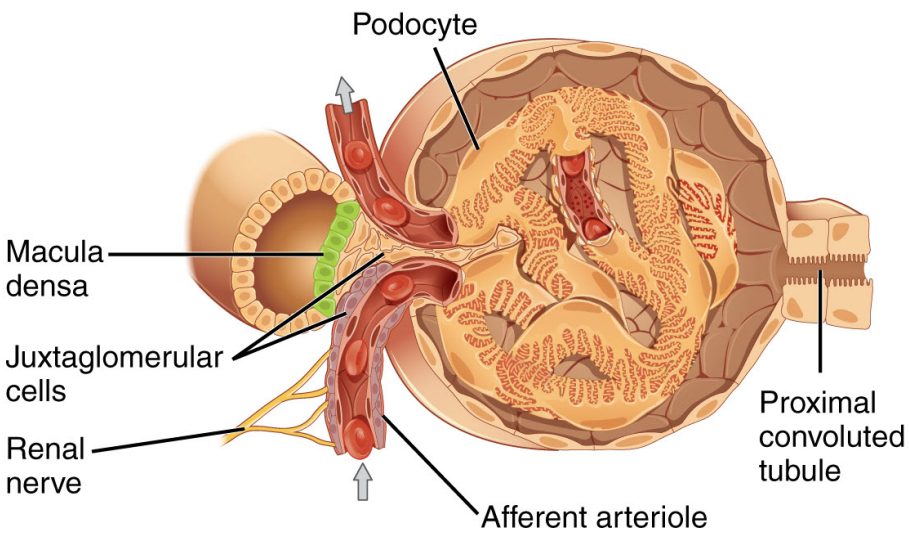


Figure 2.2: Image of a glomerulus, demonstrating the interdigitating podocytes investing the capillary loops and the location of the juxtaglomerular apparatus. OpenStax [40]

The Proximal Convoluted Tubule

The dilute urine passes through the proximal convoluted tubule (PCT) which selectively reabsorbs most of the initial filtrate, including 50%–60% of all filtered Na^+ and water. Here, metabolically demanding Na^+/ K^+–ATPases generate a trans-cellular sodium gradient to enable its recovery from the urine. Water follows by 'solvent drag' and is thus reabsorbed. The electrochemical gradient generated from active sodium transport allows other molecular pumps to recover ions (calcium and phosphate), bicarbonate ($\sim$90% of total), and up to 100% of the glucose and amino acids initially filtered. The high oxygen requirements of this part of the nephron is reflected in the large numbers of mitochondria contained within the cells of the epithelium. This also makes them particularly susceptible to hypoxia and metabolic toxins.

The Loop of Henle

Further active reabsorption of sodium, potassium and chloride takes place through the Loop of Henle, where the combination of active ionic reabsorption and watertight endothelial junctions generates a steep concentration gradient in the hypoxic renal medulla, allowing for the later reabsorption of water from the collecting duct.

There are in fact two classes of nephron in humans, and only 15% of nephrons, the *juxtamedullary* nephrons, actually project Loops of Henle deep into the hypoxic medulla in order to concentrate the urine. The remaining 85% of nephrons, the *cortical* nephrons,

project only shallow loops down to the outer medulla, and are less able to concentrate the urine.

The Distal Convoluted Tubule

The distal convoluted tubule (DCT) is the main site of urinary pH balance, excreting acids in the form of H^+ ions and the nitrogenous waste product ammonia (as the $NH_4{}^+$ ion), and reabsorbing approximately ∼90% of excreted $HCO_3{}^-$. It also actively secretes 5-10% of the total urinary creatinine (adding to that passively filtered at the glomerulus), and is the primary site of action of the body's chief calcium-recovering agent, parathyroid hormone (PTH), activating selective Ca^{2+} transport proteins on the lumenal surface of the endothelium. Urea also enters the urine at this stage by passive diffusion down its concentration gradient.

The Collecting Duct

The collecting duct then performs the final readjustment (<5%) of its ionic composition and pH (through H^+–ATPases) but more importantly the reabsorption of water via the medullary concentration gradient generated by the Loops of Henle, and in response to two key hormones – antidiuretic hormone (ADH) and aldosterone.

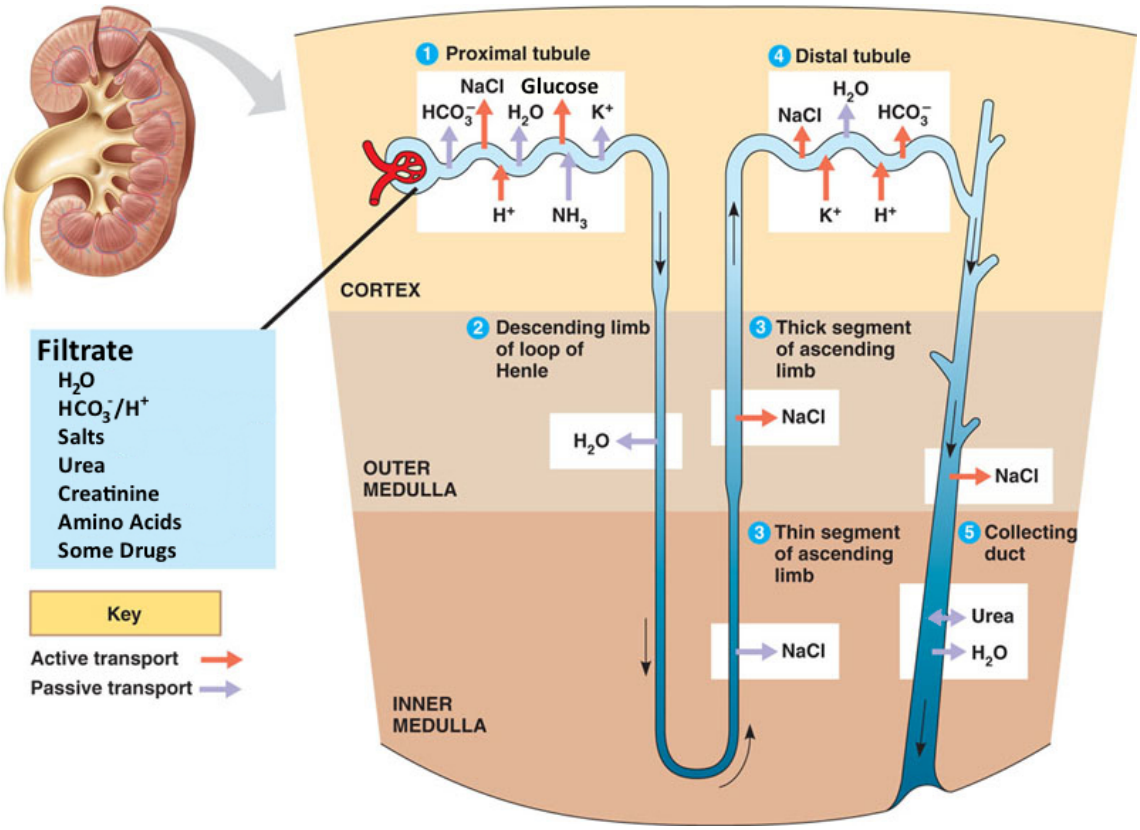


Figure 2.3: A simplified overview of ionic adjustments performed along the nephron. ©2010. Trevor Gallant, Kennebecasis Valley High School, Rothesay, Canada.

Once produced, the urine is actively peristalsed down the ureters at an average rate of 1-2ml/min[41] to enter the bladder where, in many clinical circumstances, a catheter is present to convey the urine to an external device for collection. The kidney is thus able to maintain the serum levels of key biochemicals within the tight normal ranges required for life by careful adjustment of the contents of the urine.

Composition of the Urine

The table below provides the reference ranges for the most abundant urinary solutes, and illustrates the way in which the nephron is able to concentrate certain metabolites nearly 1000–fold, particularly the nitrogenous metabolic wastes, whilst completely excluding or reabsorbing others.

NB *Each hospital laboratory has slightly different reference ranges for the normal values encountered in their patient population (95% C.I. – 2 S.D.) due to regional variations and the testing methods they employ.*

Normal Ranges in Blood and Urine [38, 41–43]			
	Constituent	Serum	Urine
General Properties	Volume	∼80 ml/ Kg Bodyweight	∼1.5 L/ 24hrs
	Osmolality	280 – 295 mOsm/ Kg	450 – 900 mOsm/ Kg
	pH	7.35 – 7.45	4.5 – 8.0
	$[H^+]$	35 – 45 nmol/L	1 – 32,000 nmol/L
	Protein	60 – 80 g/L	≤150 mg/ 24hrs
Ions	Na^+	135 – 145 mmol/L	40 – 220 mmol/24hrs
	K^+	3.5 – 5.0 mmol/L	25 – 120 mmol/24hrs
	Ca^{2+}	2.0 – 2.5 mmol/L	2.5 – 7.5 mmol/24hrs
	Mg^{2+}	0.6 – 0.8 mmol/L	3 – 4.5 mmol/24hrs
	Cl^-	95 – 105 mmol/L	100 – 250 mmol/24hrs
Nitrogenous Wastes	Urea	1.5 – 5 mmol/L	420 – 720 mmol/24hrs
	Creatinine	60 – 120 μmol/L	7 – 16 mmol/24hrs
	Ammonia/NH_4^+	10 – 35 μmol/L	20 – 70 mmol/24hrs
	Uric Acid	180 – 480 μmol/L	1.4 – 4.4 mmol/24hrs
Cells	Erythrocytes	4 – 5 x10^{12}/L	0 – 3 / HPF†
	Leukocytes	4 – 9 x10^9/L	0 – 2 / HPF†
Other	Glucose	4 – 6 mmol/L	Absent
	Bilirubin	2 – 25 μmol/L	Absent
	Ketones	Absent	Absent
	Nitrites	Absent	Absent

†/ HPF = per High Powered microscope Field

Table 2.1

However, this table is far from exhaustive – over the past half-century alone we have discovered traces of well over 100 biochemicals and metabolites in the urine which could provide insights into normal function and disease.

2.3 Renal Endocrine Function

The kidney not only has a homeostatic role in the composition and excretion of the body's salt, water and soluble wastes, but it also has a far reaching systemic endocrine role through the secretion of renal hormones.

These hormones affect the body's Ca^{2+} balance, oxygen carrying capacity and blood pressure.

Role in Ca^{2+} Homeostasis

Calcium is a tightly-regulated intracellular anion responsible for activating many metabolic and genetic cascades, as well as stabilising excitable cell membranes. Its absorption and excretion from the gut and kidney, and storage in the bones, are tightly regulated by a host of hormones. One of these is a particular form of Vitamin D - 1,25-hydroxy-Vitamin D_3.

In all mammals exposed to sunlight, the carbon skeletons of cholesterol molecules are converted to the inactive precursor of Vitamin D through interaction with UV-B radiation. In the liver this precursor, Vitamin D_3 (*cholecalciferol*), is hydroxylated at the 25th carbon to form 25-hydroxyvitamin D_3 (*calcifediol*).

Through a negative feedback loop dependent on its own product levels and other hormones involved in Ca^{2+} homeostasis, the proximal tubules of the kidney express a specific 1α-hydroxylase enzyme that acts on circulating 25-hydroxyvitamin D_3 to form 1,25-dihydroxyvitamin D_3 – the biologically active form of Vitamin D known as *calcitriol*.

This active hormone is released into the bloodstream to act on end-receptors in multiple tissues to regulate Ca^{2+} homeostasis.

Role in Oxygen Carrying Capacity

The majority of the blood's oxygen is transported by hæmoglobin molecules, which greatly increase the carrying capacity beyond that of simple solubility (page 58).

In order to reduce the osmotic pressure of large hæmoglobin macromolecules in solution, organisms with complex circulatory systems have evolved specialised hæmoglobin carrying cells which package up many millions of hæmoglobin molecules inside a single cell. In humans and other mammals these are known as *erythrocytes*, literally *red blood cells*. The red colour derives from the absorption spectrum of O_2-bound hæmoglobin.

Whereas production of erythrocytes takes place in the bone marrow from precursor stem cell populations, the actual biological oxygen sensor resides in the kidney. The communication between these two sites takes place via the hormone*erythropoetin* (EPO). Over 90% of circulating EPO derives from the kidneys, with the remaining fraction from the liver. The presence of EPO is facultative for the maturation of marrow erythrocyte stem cell lines, explaining why anæmia is common in advanced renal disease.

For reasons which were previously discussed in Section 2.2 (page 49), the convoluted tubules and ascending Loops of Henle have very high oxygen requirements in order to maintain normal membrane pump function. Populations of cells residing deep in the juxtamedullary region of the kidney, where the gradient of the O_2-hæmoglobin binding curve is at its steepest (page 64), are responsible for oxygen sensing and release of *erythropoetin*.

These cells are known as *REPOS* cells - Renal Erythropoetin and Oxygen Sensors - but the exact intracellular cascades underlying their oxygen sensing capabilities are still not yet fully understood[44].

Role in Maintaining Blood Pressure

The blood pressure in a closed-loop circulatory system is dependent on a number of factors, expressed in the following physiological 'equation':

$$\mathrm{BP} = f(\mathrm{CO}, \mathrm{BV}, \mathrm{TPR})$$

The variables in this equation are:

- BP – Blood Pressure
- CO – Cardiac Output (sometimes given as $\dot{\mathrm{Q}}$)
- BV – circulating Blood Volume
- TPR – Total Peripheral Resistance ($\propto$ vascular tone)

Blood pressure will rise with increased cardiac output, circulating volume, and total peripheral resistance. Each of these components can be targeted in order to treat hypertension[1] in human patients.

The kidney has a role in maintaining both BV and TPR. It maintains the normal circulating intravascular volume through production of the urine in response to hormonal factors, and it maintains the tone of the peripheral vasculature by releasing *renin* and Nitric Oxide (NO)

[1]High blood pressure - a very common and dangerous disease

from cells of the juxtaglomerular apparatus (JGA) which act as the kidney's pressure/flow sensor. See Figure 2.2 on page 49 for an illustration of the JGA.

The release of Nitric Oxide (NO) acts locally to increase medullary blood flow and Na^+ excretion in response to increased glomerular filtration, as a way to reduce the body's total circulating volume and thereby reduce blood pressure[45]. In this way, NO also reduces renal O_2 consumption, as Na^+ reabsorption is the major ATP-consuming process of the kidney.

Renin is an enzyme rather than a hormone. Its downstream effects include cleaving a vasoactive precursor synthesised by the liver (angiotensinogen) into its active form (angiotensin I), which is further cleaved into the highly potent vasoconstrictor angiotensin II by the endothelial cells of the lungs. Note that all of these vasoactive agents and precursors are flushed from transplant kidneys at the time of retrieval.

Angiotensin II causes peripheral vasoconstriction of all vascular beds, but in the kidney it selectively constricts post-glomerular arterioles to increase back-pressure at the glomerulus in an attempt to stimulate urine production. It also acts on the adrenal glands to increase *aldosterone* production in order to reabsorb water from the collecting ducts, thereby increasing circulating volume and blood pressure.

These cascades are part of a phenomenon known as *tubuloglomerular feedback*, whereby the kidney is able to modulate its own glomerular perfusion by sensing activity in the tubules. They are activated by two major stimuli at the JGA:

i. Fall in Arterial Blood Pressure

The JGA contains specialised pressure sensitive cells which mechanically transduce the pressure in the kidney's efferent arterioles into the suppression of renin release. If the perfusion pressure of the kidney falls, the barosensitive cells are disinhibited and release renin into the blood stream.

ii. Fall in Tubular Sodium Content

Specialised cells in the *macula densa* of the JGA are also able to sense the sodium concentration of the urine passing through the distal convoluted tubules, which acts as a surrogate of the urine production rate and thus glomerular and renal perfusion.

If the glomerular flow rate falls, and hence the filtration rate falls, the cells of the JGA will sense this as a fall in tubular [Na^+] and will attempt to restore the filtration rate by increasing the renal perfusion pressure through the release of renin from the JGA.

2.4 Blood Pressure & Tissue Perfusion

Before entering the discussion about renal metabolism, there is an important concept to be introduced – that of *tissue perfusion*.

Blood flow is understood to represent the blood volume delivered per unit time. This is classically described by Poiseuille's law, which describes the role of the vessel's length and radius, the pressure gradient across the vessel (ΔP), and the viscosity of the blood on determining the flow in the following manner:

$$\dot{\mathrm{Q}} = \left(\frac{\Delta \mathrm{P} \cdot \pi \cdot \text{Vessel Radius}^4}{8 \cdot \text{Vessel Length} \cdot \text{Blood Viscosity}} \right)$$

Here, ΔP represents the pressure gradient across an individual vessel. Whilst the Poiseuille equation cannot be generalised to describe the network within an entire organ, its variables are those most important to organ perfusion. Instead, when talking about whole organ perfusion, we use the Mean Arterial Pressure (MAP), which is the geometric mean of the systolic and diastolic blood pressure readings, where the diastolic phase is twice that of the systolic phase:

$$\mathrm{MAP} = \frac{(2 \cdot \text{Diastolic BP}) + \text{Systolic BP}}{3}$$

Perfusion is then proportional to the MAP, the aggregate tone of the blood vessels within an organ (a pseudo-sum of vessel radius) and inversely proportional to the viscosity and total vessel length within the vascular bed. A further key concept is that of matching the flow rate through an organ to the amount of tissue requiring blood:

$$\text{Perfusion} = \frac{\text{Blood Flow}}{\text{Amount of Tissue}}$$

Where the result is usually normalised per 100g of tissue.

Recalling the 'equation' of the blood pressure described previously on page 54, it is clear that a number of factors can negatively the matching of blood flow to the underlying metabolic requirements of the tissue. Some examples are given here, but this is certainly not exhaustive:

- reducing the driving pressure gradient into the organ *e.g.* by reducing the MAP through increasing the diastolic or reducing the systolic pressure
- increasing the blood viscosity *e.g.* by cooling the perfusate
- reducing the average vessel radius within an organ and thus increasing the TPR *e.g.* through normal vasomotor activity, or through pathological tubular or endothelial cell œdema compressing the outside of the vessel

Autoregulation of Tissue Perfusion

Tissues are acutely sensitive to alterations in their perfusion, and employ multiple feedback mechanisms to ensure adequate endogenous blood perfusion to meet their metabolic demand. The body's physiological systems most commonly affect control over local tissue perfusion by manipulating the cross-sectional area of vessels in the local microcirculation through the autonomic nervous system, and systemically through renal mechanisms.

Renal mechanisms includes the effects of molecules released into the circulation to:

i manipulate the renal perfusion pressure – the release of renin from the JGA of the kidney increases the peripheral vascular tone and thus the renal perfusion pressure

ii encourage fluid retention in the collecting duct to increase the circulating blood volume – the release of aldosterone from the adrenal gland as a downstream effect of renin release, and the release of antidiuretic hormone from the brain in response to baroreceptors and high blood osmolality

iii modulate the intrinsic vascular tone within the kidney to control the glomerular filtration rate – local intra-renal NO release and *purinergic* signalling

The following graph demonstrates the phenomenon of pressure-flow *autoregulation* in the kidney, and the normal perfusion pressure range over which it applies:

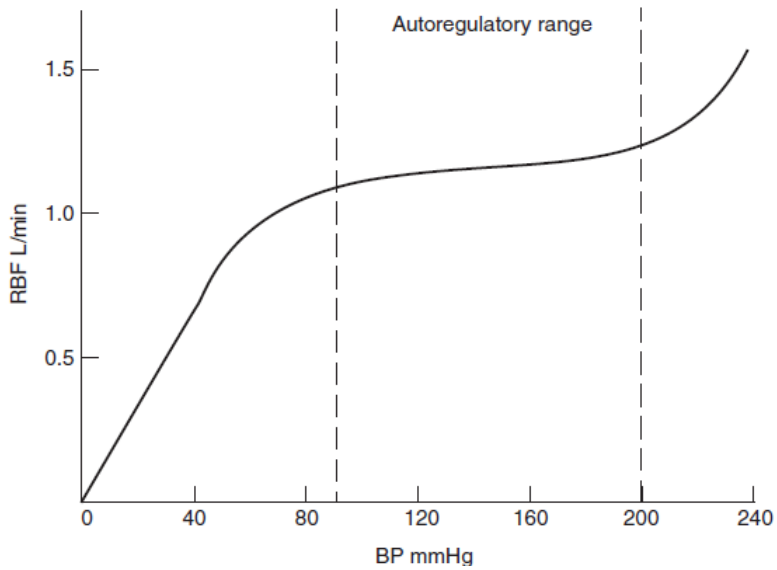


Figure 2.4: The Renal Autoregulation Curve relating Mean Arterial Pressure to Renal Blood Flow (RBF)

Note how the usual perfusion pressures applied during HMP[1] ($^{60}/_{40}$, page 38) and Brasile's WMP methodology[1] ($^{50}/_{30}$, page 41) result in mean perfusion pressures below

[1]Pressures in mmHg

the autoregulation threshold, where the renal blood flow is most sensitive to changes in perfusion pressure[25, 29].

The reason cited for preferring hypotensive perfusion over normotensive perfusion is usually tissue œdema, but there is evidence from complex human aortic vascular surgery that a kidney immediately returned to normothermic, normoxic, normotensive perfusion with a minimal period of ischæmia (<30 minutes) will function well.

Surgery on juxtarenal aortic aneurysms requires reconstruction of one or both renal arteries. This involves clamping, reconstructing, and unclamping the renal artery – the closest paradigm we have to human renal autotransplantation with a brief period of ischæmia. In the most recent case series of 200 patients, only 11% experienced a brief period of acute renal injury, and only 0.5% ended up with renal failure as indicated by the need for permanent hæmodialysis[46].

There do not appear to be any major papers specifically focussing on autoregulation under HMP conditions, and whether perfusion pressures below the autoregulation range are harmful to the parenchyma in itself, for example through collapse of the *vasa rectæ* capillaries and vascular shunting. However, there is human evidence that higher renal perfusion pressures within the autoregulatory range leads to better renal oxygenation and filtration function in the setting of acute renal injury[47].

2.5 Renal Metabolism

Designing devices to mimic normal physiology requires a discussion of our current state of knowledge of renal oxygen delivery, utilisation and cellular metabolism in order to meet their normal metabolic demands.

Renal Oxygen Delivery

Oxygen is the fundamental enabler of animal life, acting as a common electron-acceptor in the final step of mitochondrial oxidative phosphorylation to regenerate energy substrates for cellular metabolism.

The evolution of the circulatory system and specialised oxygen transporting molecules some 600 million years ago were fundamental to the development of macroscopic life, allowing organisms to meet their metabolic requirements at rates greater than those allowed by diffusion of oxygen down a concentration gradient, or simply by oxygen in solution[48]. For example, the mammalian oxygen transporting molecule hæmoglobin allows the blood to carry around 6,000 times more oxygen than the amount dissolved in solution.

Tissue oxygen delivery is a function not only of the*amount* of oxygen carried per volume of blood, but also the *rate* of blood supply to that tissue.

Blood Oxygen Content

The key oxygen transport molecule in mammals is hæmoglobin A (HbA), a tetramer comprising 2 subunits of β-hæmoglobin and 2 subunits of α-hæmoglobin. This forms over 97% of all hæmoglobin in adult red cells (erythrocytes), the remaining 3% being made up of hæmoglobin A_2 or other subtypes. Within each α and β subunit is a hæme molecule containing a central ferrous Fe^{2+} ion which coordinates with oxygen. On doing so, the 3D conformation of the globin subunit surrounding the hæme molecule changes, leading to conformational changes of the other globin subunits which favour oxygen binding. Thus O_2 binding and unbinding is cooperative in nature.

The following binding curve is found in all physiology textbooks and illustrates the affinity of hæmoglobin for oxygen in the local environment, given as the percent of occupied O_2 binding sites for a given partial pressure of O_2 (pO_2). The percent O_2 site occupancy is known as the *oxygen saturation* (sO_2) of the blood. In this graph, the right hand side provides a convenient indication of the amount of O_2 carried in ml/100ml of blood at a average adult male hæmoglobin concentration of 150g/L.

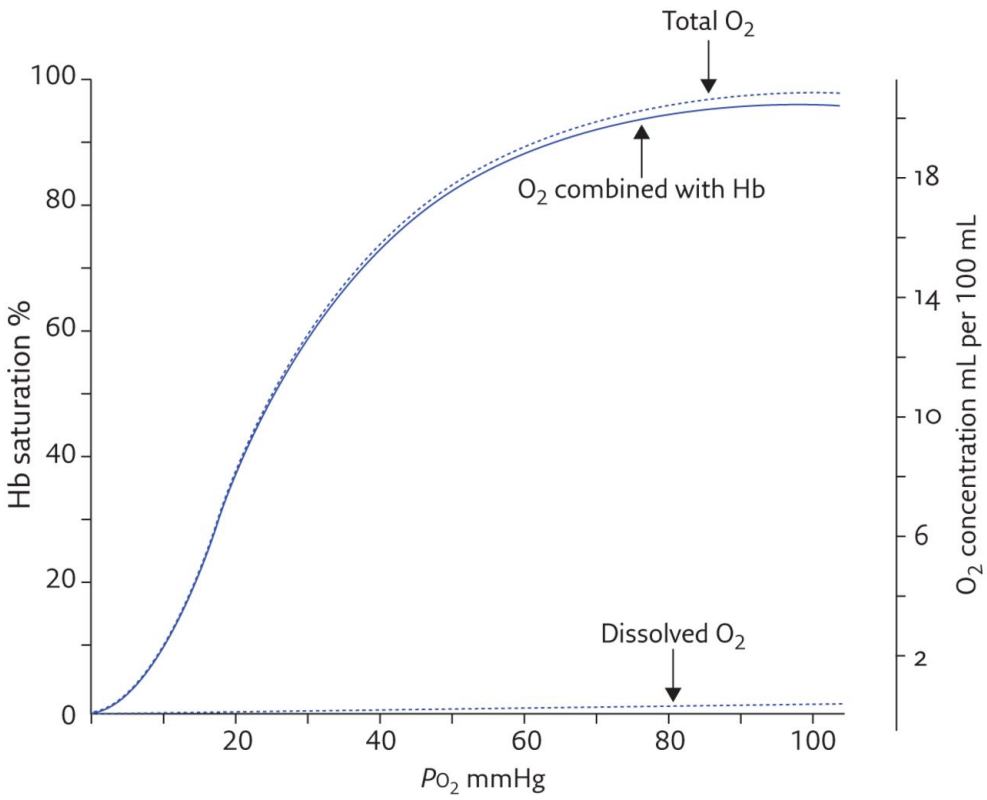


Figure 2.5: The sigmoidal O_2-hæmoglobin binding curve [49]

Many local factors affect the slope and mid-point of this sigmoidal binding curve, including:

- Temperature
- O_2 and CO_2 tension
- pH
- 2,3-bisphosphoglycerate (2,3-BPG) – a byproduct of cellular respiration

In hot, hypoxic, acidic conditions with larger concentrations of 2,3-DPG, O_2-offloading is favoured and the curve shifts to the right, and vice-versa in well-oxgenated or slowly-metabolising environments. The subunit mix comprising specific hæmoglobins determines their oxygen binding affinity and responsiveness to these various factors under different ambient conditions, and thus the shape of the O_2-binding curve.

Relating the blood content of hæmoglobin to the oxygen carrying capacity of blood is the experimentally-determined *Hüfner's constant* of 1.34mlO_2/g Hb. In addition to this, a small fraction of oxygen is transported in solution in the plasma (0.03ml/L/mmHg pO_2).

The total volume of oxygen carried in each litre of arterial blood for a normal adult male human at sea level is therefore:

$$\begin{aligned} \mathrm{vO_2} &= (1.34 \cdot [\mathrm{Hb}] \cdot \mathrm{sO_2}) + (0.03 \cdot \mathrm{pO_2}) \\ &= (1.34 \cdot 150\mathrm{g/L} \cdot 100\%) + (0.03 \cdot 100\mathrm{mmHg}) \\ &= (201\ \mathrm{ml}) + (3\ \mathrm{ml}) \\ &= 204\ \mathrm{ml\ O_2\ /\ \ Litre\ of\ blood} \end{aligned}$$

Severinghaus and Roughton determined that cooling human blood from from 37°C to 0°C increased the O_2 affinity of hæmoglobin some 22-fold, but only increased the water solubility of O_2 by 2-fold[50]. Therefore under a hypothermic blood perfusion paradigm there will be slightly more oxygen carried in solution, but the total oxygen delivery will be far lower due to a left shift in the hæmoglobin binding curve, and a reduction in the diffusion constant for oxygen in free solution from 3.45×10^{-5}cm^2/s at 37°C to 1.10×10^{-5}cm^2/s at 0°C[51].

Renal Blood Flow

The kidneys of a healthy adult human receive some 20% of the total cardiac output, between 1,000ml/min – 1,200ml/min. Each kidney therefore receives half of this supply, with the caveat that not all kidneys are identically paired in size or function, but are usually around 150g in weight. The kidney also employs various autoregulation strategies to maintain constant perfusion at normal arterial pressures (page 57).

Of this total blood flow, the cortex of each kidney receives around 4.2ml of blood/minute/gram of tissue, and the medulla only 1.9ml/min/g [52]. This difference in regional blood perfusion derives from the unique capillary anatomy of the renal medulla and actually underlies the kidney's ability to concentrate urine (see page 61 for further discussion).

The total amount of oxygen delivered to each kidney is therefore in the region of 204ml O_2/L of blood $\times$ 500ml/minute blood flow, or $\simeq$100ml O_2/minute.

Renal Oxygen Usage

The active transport processes of the kidney are highly energetic, and consume large amounts of oxygen and metabolic substrates (glucose, lactate and amino acids) in order to 'pump' selected ions and macromolecules back against their concentration gradients. These transport processes are often based on Na^+ co-transport or exchangers.

The majority of this highly energy-intensive active transport takes place in the convoluted tubules and ascending Loops of Henle which are located in the renal cortex and outer medulla where the capillary network is best able to maintain an O_2 supply. These cell populations contain large numbers of mitochondria to regenerate the phosphonucleoside molecules (ATP and GTP[53]) required to power the active membrane pumps, and are accordingly highly sensitive to changes in local O_2 supply.

Passive transport processes of urea and water predominate in the medulla, via the descending limbs of the Loops of Henle and the collecting duct. This hypoxic environment is the byproduct of the requirements of the concentrating mechanism for urine. Rich capillary beds here would reduce the osmotic gradients of urea and other substrates required to draw water from the collecting ducts. Instead, the unique anatomy of the renal medulla has multiple small capillary vessels plunging deep into the medulla from the arterioles of the cortex, forming tight hairpin loops of slow-flowing capillaries, the *vasa rectæ*, investing the space around each Loop of Henle (see Figure 2.1 on page 48 for an illustration of these vessels).

The low flow rate, the unique anatomy, and endothelial transport proteins found in these vessels allows for oxygen and substrate delivery without washing out the metabolites required to maintain the local osmotic pressure.

As a result, the oxygen tension (pO_2) of the renal cortex lies around 50mmHg, and the medulla only some 10mmHg – 20mmHg[52]. For comparison, the partial pressure of arterial blood in a healthy human at sea level is around 100mmHg.

The cell populations within the renal medulla and juxtamedullary region are therefore more

dependent on anærobic glycolysis for ATP generation, but they are also more susceptible to hypoperfusion [52] than cortical cells, which are well-supplied with capillary blood.

This explains the selective insults seen in response to ischæmia–reperfusion injuries, preferentially affecting the energetic proximal tubules and ascending limb of the Loop of Henle[54], followed by the less susceptible descending Loop of Henle and the distal tubule, and finally the collecting ducts and glomeruli which are affected the least.

Oxygen Consumption Increases with Perfusion and Sodium Load

Unlike other organs, the delivery of blood to the kidney actually drives its oxygen demand rather than acting in response to it.

Increased blood perfusion leads to an increased glomerular filtration rate. This places additional metabolic demand on the active transporters in the cortex and juxtamedullary regions of the kidney to reabsorb and rebalance the additional load of ions and substrates in the urine.

Sodium transport in particular drives oxygen consumption, primarily because the transport of Na^+ ions drives most of the other solute gradients. Indeed, Na^+ transport alone accounts for some 70% of renal oxygen demand, as established by early isolated cyanide infusion experiments in dogs and other experiments[54, 55].

Renal Oxygen Consumption Experiments

One of the earliest experiments to investigate human renal oxygen consumption is that of Walter H. Cargill's group in 1949[56]. In an ingenious setup, they were able to directly sample blood from the renal veins of willing volunteers with simultaneous arterial blood oxygen determination.

Cargill's results showed that as well as receiving a substantial blood flow and perfusion rate, a pair of normal kidneys extract 16mlO_2/min ($\pm$2.9mlO_2/min) from a mean blood supply of 1155ml/min ($\pm$229ml/min).

The same experiment also demonstrated that this *extraction ratio* of 10%–16% v/v (5% – 8% per kidney) remains constant in renal disease, where the kidneys of volunteers with chronic nephritis only extracted 6.1mlO_2/min ($\pm$3.3mlO_2/min) but from an equally reduced renal blood supply of 353ml/min ($\pm$176ml/min).

A more recent repetition of Cargill's experiment was undertaken by Dr. Brenda R.C. Kurnik *et al* in 1992. In their study, 60 patients undergoing elective cardiac catheterisation also underwent femoral arterial oxygen sampling, renal venous oxygen sampling, determination

of renal blood flow by thermodilution, and submitted 12-hr overnight urine collections to allow calculation of their baseline renal function (see Section 2.7.1 on page 70 for further discussion of this aspect).

Kurnik's experiment determined that normal kidneys extract 8.76mlO_2min/kidney $\pm$2.26mlO_2 each[1], from a mean renal blood flow of 527ml/min/kidney ($\pm$146ml/min).

And so while the kidneys only account for less than 0.5% of the total body weight, each one consumes nearly 8% of all O_2 being delivered.

However, compared to other well-perfused organs like the heart which extracts nearly 45%[57] of the O_2 it receives, the kidney has a large excess oxygen supply, explaining how it is able to 'ramp-up' its O_2-demanding functions when required, for example to match an increase in the tubular ultrafiltrate delivery following an increase in perfusion during exercise.

Meeting this Demand

It should be clear that meeting the 12–18mlO_2/min O_2 demand of a normally functioning human kidney will not be possible through O_2 in solution alone, where O_2 solubility is $\simeq$0.03mlO_2/mmHg pO_2.

This explains why the historic trend of hypothermic preservation took hold so strongly. Cutting down the renal metabolic rate, reducing its requirement to extract Na^+ by flushing with solutions that match the intracellular composition, and then providing zero glomerular filtration are all ways to theoretically reduce renal O_2 consumption.

However, this is no longer good enough. Recent experiments[58] show that even hypothermic kidneys perfused with acellular media at 8°C continue to consume O_2 at rates sufficient to exhaust the supply in solution under normoxic conditions (150mmHg pO_2, delivering 4.5mlO_2/min).

Calculating O_2 Consumption

It is possible to calculate the amount of O_2 extracted from the blood as a function of the difference between the arterial (s_AO_2) and venous (s_VO_2) blood oxygen saturations, and the hæmoglobin concentration if we assume minimal losses into the parenchyma through bleeding or œdema and thus 100% arterio-venous flow concordance.

$$\text{Arterio–Venous } O_2 \text{ Extraction} = (1.34 \cdot [\text{Hb}]) \cdot (s_AO_2 - s_VO_2)$$

[1]Converted from the data within Kurnik's paper using their value of 44.6μmol O_2/ml O_2

2.6 Renal Energy Substrates

The kidney is able to extract energy from multiple substrates delivered in the blood including fatty acids, glucose, lactate, glutamine, citrate, ketone bodies and amino acids in the following approximate proportions[55]:

- Glucose – 13%
- Glutamine – 35%
- Lactate – 20%
- Free Fatty Acids – 15%
- Citrate – 7%

Most of our evidence for renal substrate metabolism derives from non-human animals and tissue sections[55, 59, 60]. As such, the applicability of these studies to whole-organ or human renal metabolism is not always entirely clear. The following illustration attempts to clarify the major differences in segmental cellular metabolism along the nephron:

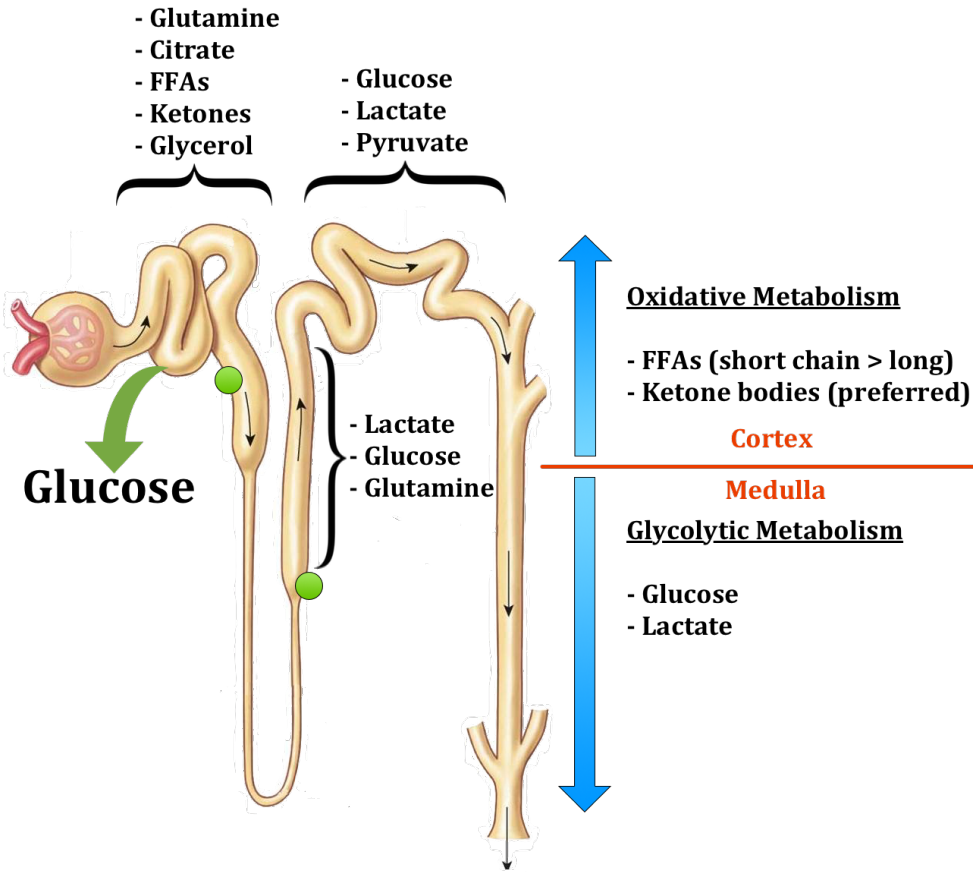


Figure 2.6: Summary of renal substrate preferences within the nephron. Green circles = glucose transporters. Précis of [55, 59]

Renal Cortex

The primary energy source for the renal cortex is the oxidative metabolism of free fatty acids (FFAs), preferring short chained FFAs such as butyric acid over longer molecules such as oleic acid. Fatty acid metabolism also appears to decrease lactate utilisation as an energy source, directing lactate instead towards gluconeogenic pathways.

However, renal cortical cells will preferentially consume ketone bodies (α-ketoglutarate or β-hydroxybutyrate) over FFAs.

The renal cortex, specifically the PCT, plays a significant role in whole-body gluconeogenesis. It is able to generate up to 20% of the body's basal glucose supply in starvation. The substrates for this oxygen-consuming process include lactate and pyruvate derived from medullary glycolysis, as well as glutamine, citrate and glycerol derived from the tubular fluid and the circulation.

The PCT is also the primary site of glucose reuptake from the urine, via the facilitative GLUT1 & GLUT2 transporters, as well as an energetic Na^+-Glucose co-transporter.

The shuttling of lactate and other anaerobic byproducts from the medulla to the cortex, with gluconeogenesis in return, accounts for the interesting observation that there is no net arterio-venous lactate difference across the kidney.

Renal Medulla

Studies on various mammal species demonstrate that the less-well oxygenated regions of the nephron, including the medullary ascending Loops of Henle and collecting ducts, and the juxtamedullary distal convoluted tubules, contain high levels of glycolytic enzymes. The oxygen tension in the medulla is not so low that some oxidative metabolism cannot take place within the mitochondria, for example the complete consumption of lactate to CO_2 or the utilisation of FFAs or glutamine.

However, for the majority of cells in the deeper medulla, glycolysis predominates.

Renal medullary cells also contain very low glycogen reserves, indicating that most of the glucose required by the medulla must be delivered via the circulation. There is also an insulin-responsive glucose transporter (GLUT4) in the ascending limb of the Loop of Henle.

Not only are these deeper regions able to consume glucose delivered to the kidney, it is also known that the renal cortex is an active site of gluconeogenesis. Thus even with substantial basal glucose consumption by the renal medulla, the net renal flux of glucose is near parity.

2.7 Renal Function Monitoring

Referring back to the discussion at the beginning of this Chapter where the roles of the individual parts of the nephron were discussed, the precise meaning of 'renal function' is not entirely obvious. It appears likely that in the future we will test for chemicals in the blood or urine to indicate injury to specific parts of the nephron so as to design targeted therapies.

However, at present the clinical use of the term 'renal function' refers to the far more simple concept of *Glomerular Filtration Rate* (GFR), discussed later in this section on page 68

Emerging Functional Biomarkers

Perhaps the most exciting group of markers to emerge in recent years are those which seem to indicate damage to specific parts of the kidney, illustrated in Figure 2.7 below.

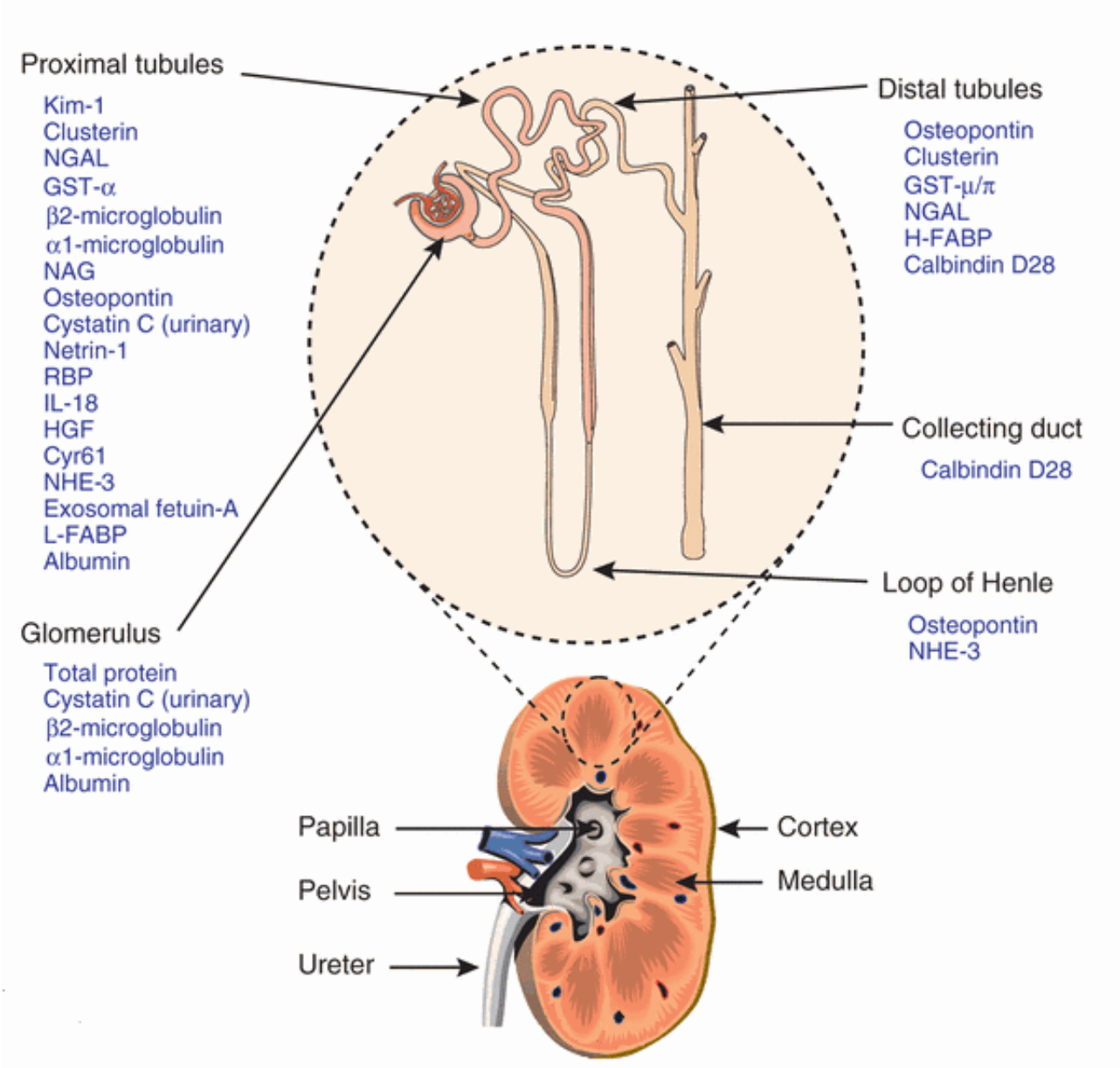


Figure 2.7: Emerging biomarkers of renal injury. From Bonventre *et al.*, 2010[61]

Specific parts of the nephron produce proteins with antimicrobial, signalling or as-yet-unknown properties. Some of these proteins are released constitutively, others in response to insult or injury causing cellular damage and 'leakiness' of their luminal plasma membranes, or even cell death.

The markers of key interest in ongoing studies are serum Cystatin C, Interleukin-18, Kidney Injury Molecule-1 (KIM-1), and in particular urinary or plasma Neutrophil Gelatinase-Associated Lipocalin (NGAL). [62–66] Elevated NGAL has been demonstrated to predict unfavourable outcomes following renal transplantation [64], after cardiopulmonary bypass [65, 67] and in cases of septic renal injury [68].

2.7.1 Glomerular Filtration Rate

The *Glomerular Filtration Rate* (GFR) is an attempt to measure the aggregate ability of all nephrons to filter the blood in order to begin the production of urine. It is often reported as a value in ml/minute normalised to a body surface area of 1.73m^2. The normal adult range of GFR lies between 90ml/min/1.73m^2 and 130ml/min/1.73m^2. Worsening GFR is the clinical means of assessing the *stage* of a patient's chronic kidney disease, where a GFR of $\leq$15ml/min is termed *end-stage renal failure* (ESRF).

Before any of the specialised emerging markers were known, or indeed much of the details of renal physiology, discussions were ongoing about the relative contributions of secretion vs. filtration/reabsorption to the overall function of the kidney. In 1926, Poul Rehberg (1895 – 1989) surmised that it would be possible to calculate the rate of filtration at the glomerulus if there was a substance that was purely filtered and not reabsorbed during its passage through the tubules[69].

It was known in his day that the chemical *creatinine* demonstrated a very high difference in concentration between the urine and the plasma, and so through the classical (and now frowned-upon) method of self-experimentation, he sought to estimate his own renal filtration rate by ingesting large quantities of creatinine by mouth and collecting serial blood and urine samples for assessment.

His equation was identical to that used today, using the fractional excretion of a substance into the urine and the rate of urine production:

$$\text{GFR (ml/min)} = \text{Urine Flow Rate} \cdot \left(\frac{[\text{Urine}]}{[\text{Plasma}]} \right)$$

His experiment found his GFR lay between 100ml/min – 200ml/min, but given that we know nothing of his height or build, it is difficult to normalise these measurements into the modern form per 1.73m^2. He also correctly surmised that the total blood flow to the kidneys must be in the region of 1.1L/min, and that filtration/reabsorption is the major function of the kidney, with a possible small contribution from secretion, and not vice versa as his contemporaries had proposed (nor indeed earlier theories that the glomerulus acted as a spinning pump to drive urine through the tubules).

Unfortunately, Rehberg's supposition that a small amount of creatinine could be secreted is true – approximately 10% of urinary creatinine is actively secreted by the distal tubules. The total amount secreted also remains constant in the face of declining renal function, and so secreted creatinine contributes a greater proportion of urinary creatinine over the filtered proportion in renal failure.

Instead, other substances have been discovered and put into clinical use which are only

filtered at the glomerulus.

However, the caveat here is that GFR studies using these purely-filtered substances *do not agree* with each other, nor do they acurately predict mobidity and mortality outcomes of patients with renal disease more accurately than the current clinical practice of estimating GFR from serum creatinine levels[70]. The use of serum creatinine concentration as a marker of renal function is discussed in the following section on page 70.

Because modern healthcare systems are called upon to perform thousands of renal function tests per day, it is also clear that the various cumbersome methods used to measure GFR would grind modern medical laboratories to a halt, and so they are reserved for use in specific circumstances.

Inulin

Inulin is a plant-derived polysaccharide which first found use in experiments into renal physiology in the 1930s. It was found that unlike creatinine, inulin is solely filtered at the glomerulus with no secretion or reabsorption in the tubules, nor is it susceptible to enzymatic digestion in the human body.

Inulin clearance therefore reflects the true GFR[71], but the administration and testing procedure has barely changed in 80 years, requiring serial and simultaneous blood and urine collections with a continuous background intravenous infusion of inulin.

Radioisotopes

There are a number of other chemical agents which are purely filtered by the glomerulus. Radiolabelled versions of these agents are used for 'single-shot' GFR measurements, whereby a single dose of the radiolabelled chemical is given intravenously, followed by simpler gamma camera or SPECT monitoring of the time course of the agent through the kidneys, versus the continuous infusion protocol required for inulin clearance measurements.

Radiotracers include ^{125}I-Iothalamate, ^{125}I-Iohexol and ^{99m}Tc-DTPA[1].

para-Aminohippuric Acid (PAH)

PAH is an exogenous chemical that is soluble in the plasma and excreted in the urine through a combination of glomerular filtration (30%) and tubular secretion (70%).

[1] diethylenetriaminepentacetate

The clearance of PAH therefore reflects the renal plasma flow (where this is equal to renal blood flow minus the fraction representing the hæmatocrit) rather than the GFR, with the caveat that PAH clearance actually underestimates the renal plasma flow by a few percent because not all renal blood supply is delivered to the glomeruli or tubules. It is not widely used in clinical practice.

PAH clearance has been superseded in clinical practice by radiolabelled ^{99m}Tc-MAG3[1] scanning.

Creatinine Clearance (C_{Cr})

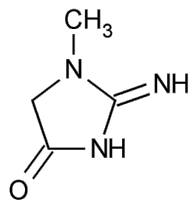


Figure 2.8: The molecular structure of creatinine

Clearance is the term used to describe the volume of blood from which 100% of a substance can be removed per unit time, thus resulting in the same units as the GFR (ml/min) but with a different underlying concept. It could be reasonably claimed that Rehberg began the practice of measuring *creatinine clearance* as a marker of renal function in his 1926 paper[69].

Creatinine is present in human blood in micromolar concentrations, nearly unmeasurable in Rehberg's day, because of constant filtration by the kidney. In a steady state system, the body's skeletal muscles will release a constant amount of creatinine into the bloodstream, and the kidneys will remove this from the circulation through a combination of filtration and active tubular secretion. This active tubular secretion comprises a larger fraction of creatinine clearance at the lower functional extreme, and leads to overestimation of the GFR. Figure 2.9 demonstrates this steady state between production by the muscles and excretion in the urine.

[1] mercaptoacetyltriglycine

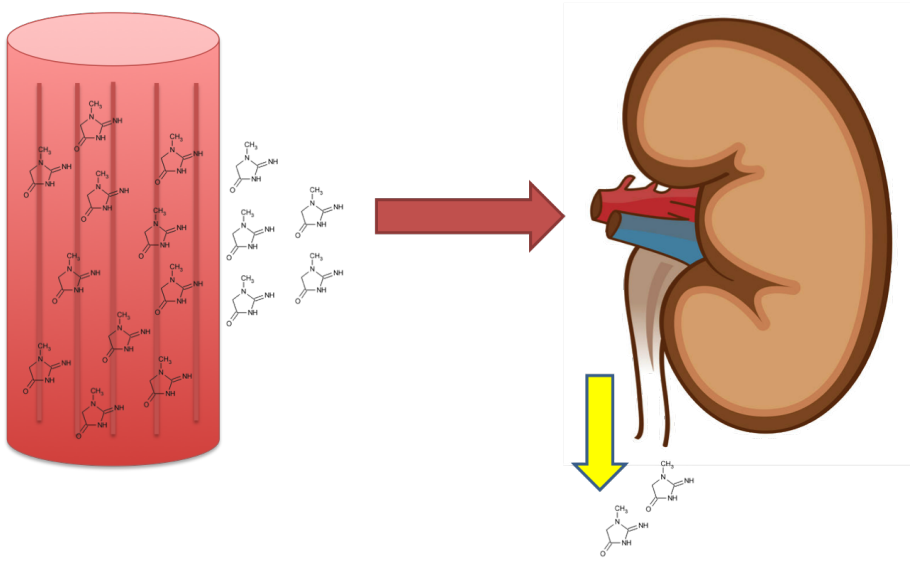


Figure 2.9: Steady-state creatinine production by skeletal muscles and excretion in the urine

Assuming a steady-state system allows us to avoid urine collection and serial measurements of plasma and urine creatinine concentrations. Instead, we are now able to take a single spot measurement of the serum creatinine concentration and relate this to the underlying renal function using equations derived from normograms of large scale human studies.

The first large scale study was undertaken by was Cockcroft and Gault in 1976 using paired 24hr urinary creatinine measurements from 249 male patients, and related the underlying creatinine clearance to age and serum creatinine measurement[72]. They also added an arbitrary 15% reduction for females, even though none were included in their study. Their formula was as follows:

$$C_{Cr} = \left(\frac{(140 \text{ - age}) \times (\text{weight in kg}) \times 1.23}{[\text{Creatinine}]\ \mu\text{mol/L}} \right) [\ \times 0.85 \text{ if female}]$$

This was then followed by the more extensive *Modification of Diet in Renal Disease* (MDRD) study which published its results in 1994, comparing ^{125}I-Iothalamate measurements of GFR against serum creatinine concentration in 1628 patients from a wider cross-section of the population with chronic renal disease[73].

The formula which arose from this study is still widely used as the standard for estimating GFR from a single measurement of serum creatinine concentration (in μmol/L), despite systematic underestimation of measured GFR in patients with better renal function, and lack of correction for patient body weight. Formula 2.1 below has been updated to reflect creatinine measurements standardised with isotope dilution mass spectrometry[74] (see page 322 in Chapter 10 for more details).

$$\begin{aligned}\text{eGFR (ml/min/1.73m}^2) &= 30849 \times \text{Creatinine } (\mu\text{mol/L})^{-1.154} \times \text{Age}^{-0.203} \\ &\times [0.742 \text{ if Female}] \\ &\times [1.212 \text{ if Black}]\end{aligned} \tag{2.1}$$

GFR measurements translate into degrees of renal dysfunction, allowing the classification of a patient into one of five categories of Chronic Kidney Disease (CKD) as shown in Table 2.2 below.

Stage	1	2	3	4	5
GFR (ml/min/1.73m^2)	>90	60 – 89	30 – 59	15 – 29	<15

Table 2.2: The 5 stages of CKD. Stages 1 and 2 have preserved function but with evidence of renal disease, such as scarring or the presence of protein or blood in the urine. Stage 5 is also known as End-Stage Renal Disease (ESRD), requiring dialysis or transplantation

Newer Formulæ

Newer formulæ for estimating the GFR and clinical outcomes from simpler clinical blood tests are always under development. Additions include blood urea and albumin levels and newer markers such as Cystatin C.

Despite all of these new developments, the standard clinical measure for renal function remains the serum creatinine concentration.

2.8 Research Themes Discussed in this Thesis

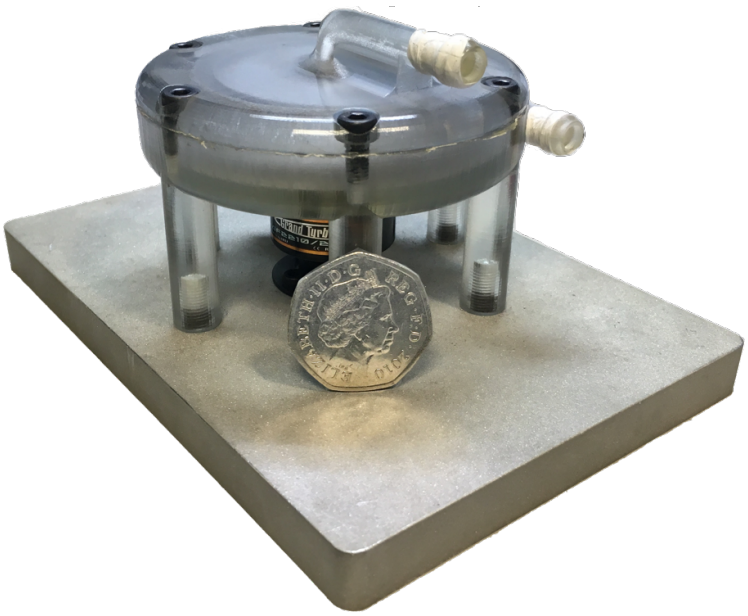
The remaining part of this Thesis is divided into three major research themes –

Part 1 (pg. 73) Designing and implementing a Tesla-type pump to create a perfusion system with minimal hæmolysis

Part 2 (pg. 155) Designing and implementing a miniature integrated sensor system for key physiological parameters of organs undergoing normothermic perfusion

Part 3 (pg. 319) Designing and implementing a live renal function monitor based on real-time creatinine detection

Part 1:
Design and Implementation of a Tesla-Type Pump



Chapter 3

The Development of Specialised Blood Pumps

Multiple technologies exist for pumping a variety of liquids, yet none have been tailored to the specific requirements of a single isolated kidney undergoing normothermic blood perfusion.

This Chapter considers the issue of biocompatibility of blood pumps and explores the range and specifications of pumping technologies with particular attention to those solutions already in use for blood pumping such as cardiopulmonary bypass pumps or Ventricular Assist Devices (VADs).

3.1 The Heart as the Pre-Eminent Biological Pump

The heart has evolved over hundreds of millions of years from a simple peristaltic tube into the multi-chambered specialised organ that powers the circulatory system of all large mammals. It operates ceaselessly for a hundred years or more (in the case of giant tortoises), has a low energy requirement (≈1.2W[1] at 20% efficiency[75], totalling 6W), has the ability to restructure itself to meet long-term changes in pressure/flow requirements, and is able to (tautologically) meet the pressure and flow needs of biological systems.

Perhaps most importantly, the heart has the ability to continuously recirculate blood *without damaging it* and without allowing clots or micro-emboli to form.

3.1.1 Basic Cardiac Physiology

The healthy adult human male contains a blood volume of approximately 5 litres for the 'standard' 75kg human. This is circulated around the body on average once per minute, with a resting heart rate of around 60 – 70 bpm[76].

Each ejection stroke of the heart ejects approximately 70mls per beat, with matching volumes passing through the left (systemic) and right (pulmonary) sides of the heart. Note that the ventricles do not fully empty with each stroke – only some $^{2}/_{3}$ of the ventricular volume is ejected with each contraction, termed the *ejection fraction* of the heart.

The left side of the heart therefore delivers 5L/minute of blood to the systemic side of the circulation, at an ejection pressure of 120mmHg (systolic pressure) in healthy adults. The Windkessel effect[2] results in the diastolic pressure of 80mmHg, hence the commonly understood '$^{120}/_{80}$' representation of a normal blood pressure. The right side of the heart delivers the same volume into the lungs, but at a systolic pressure of 25mmHg and a diastolic pressure of 8mmHg.

Any pump intending to replace the function of the heart must therefore be able to match these pressure and flow values, with some leeway for safety. For example, the pumps used for extracorporeal cardiac bypass circuits must be capable of generating laminar flows of 7L/min with pressure heads up to 7× higher than native pressures due to the need to counteract the pressure drop along the tubing to and from the patient[77]. Section 3.2.3 on page 82 explains why the blood flow must remain laminar.

Pumps acting as *ventricular assist devices* (VADs) are in closer proximity to the circulation they are supporting and as such their pressure requirements will be lower to match those

[1] Simplifying: 6L/min = 100ml/sec. Density ≈1g/cm^3. 90mmHg mean pressure = 12000 Pa
[2] Recoil of the elastic tissue within the walls of all major vessels

of the native circulation. The flow requirements may also be lower than required for total cardiac replacement because there will be some residual left ventricular function, even if the organ is failing.

In the case of pumps for isolated perfused organs, the pressure will need to be able to replicate normal systemic values, but the volume flow need only match that for the single organ. Experiments with normothermic single kidney perfusion have to date used commercial-grade pædiatric cardiopulmonary bypass machines, with their concomitant lower pressure and flow rates than adult devices[33, 34]. The requirements of an isolated perfused kidney were covered in more detail in Section 2.5 on page 60.

3.2 Biocompatibility, Hæmolysis and Thromboembolisation

There are two key aspects to consider when discussing the biocompatibility of a medical device. These are (i) the biocompatibility of its *function* and (ii) the biocompatibility of its *behaviour*.

The first of these, the biocompatibility of its *function*, is the matching of the pump's pressure/flow output to that required by a biological system, and a large number of blood pumps exist which are able to match the functional requirements of the human circulatory system. These will be covered in depth in Section 3.4, from page 93 onwards.

The other aspect of biocompatibility is the *behaviour* of a pump in a biological environment.

Ensuring the exterior housing or connection tubing of a pump does not engender infection, sepsis, excessive scarring or prevention of other native biological function[1] is relatively straightforward and depends upon the materials chosen, the sterility of the manufacturing process and the skill of the surgeon.

What is less straightforward to model or predict is the biocompatibility of the internal working surfaces of a blood pump. Blood is not a simple liquid, but rather comprises multiple cellular and acellular components which can be damaged. Blood also behaves in a non-Newtonian manner under motion or acceleration[2], and the study of blood rheology has been a topic of great importance for many decades[78].

In this Section I will consider the susceptibility of blood and its components to damage, namely *hæmolysis* – the 'splitting' of red blood cells as a result of membrane rupture.

[1] Such as digestion in the case of VADs implanted in the abdomen
[2] Blood viscosity is not constant and changes according to shear rate

Some degree of hæmolysis is natural in the human circulatory system as a result of senesecent red cells no longer able to maintain the integrity of their membranes, or it can occur pathologically in the case of genetic abnormalities in red cell conformation or abrupt changes in flow which exist in the cases of damaged heart valves or atherosclerotic arteries.

I will also briefly consider blood coagulation and the components involved in the clotting cascade to cause thrombembolisation.

Understanding and counteracting these two factors – hæmolysis and thromboembolisation – are perhaps the greatest challenge facing engineers of long-term implantable blood pumps such as VADs.

In order to prevent hæmolysis and thromboembolism in isolated organ perfusion, attempts have been made to replace the blood with macromolecular polymers of hæmoglobin[29] or other oxygen transporters such as perfluorocarbons[26], but these are not without problems of their own (see page 37 for a discussion of the issues).

The pump design proposed in this Thesis has been tailored to an erythrocyte-based perfusate which has been modified to remove clotting proteins. The reasons for this are discussed later on page 122.

We will begin by exploring the role and structure of erythrocytes in the blood in order to come to an understanding of conditions which may lead to hæmolysis.

3.2.1 Erythrocyte Biology

Discussed already to some degree in Section 2.3 in the Chapter on Renal Physiology (page 53), erythrocytes comprise some 40% – 50% of the blood by volume[1] and specialise in the transportation of oxygen around the circulatory system using hæmoglobin as a carrier molecule. Simply transporting O_2 through freely-circulating hæmoglobin molecules would exert tremendous osmotic pressure on the biological membranes in contact with the blood.

Erythrocytes reduce this osmotic pressure by tightly packaging around 250,000,000 individual hæmoglobin molecules into the equivalent of one macro-crystal contained within a membrane, the red blood cell. In order to maximise this packing density, mammalian erythrocytes have evolved to eject all organelles during the final stage of their development in the bone marrow including ribosomes, mitochondria and nuclei .

As a result, mature circulating erythrocytes are unable to perform any protein synthesis. Instead they come pre-packaged with all the proteins on which they rely to maintain their cytoskeletons, a variety of pumps and transporters to maintain their membrane osmotic

[1]The *hæmatocrit* - equalling between 120g/L – 150g/L in healthy adults

pressure and electrolyte balance, and the enzymes required for ATP generation through anærobic glycolysis.

Conformational Changes

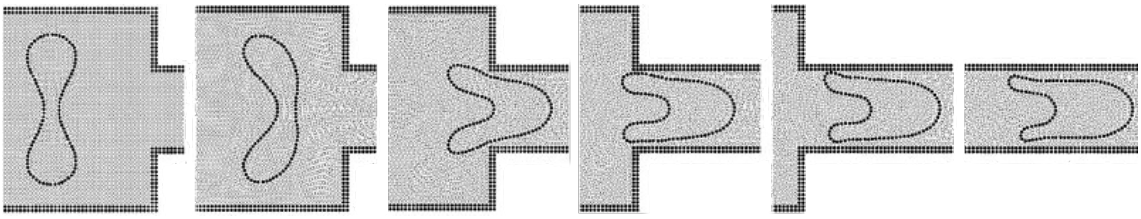


Figure 3.1: Simulated erythrocyte deformation in a thin channel demonstrating transition from a biconcave disc in bulk circulation into the 'parachute' conformation required for narrow capillary channels. From Hosseini *et al.*, 2009[79]

Human erythrocytes must be able to alter their conformation as they traverse the circulatory system. In larger vessels and bulk flow, erythrocytes adopt a 'biconcave disc' conformation to minimise surface energy, measuring some 8μm in diameter and 2.2μm in thickness around the rim but only 0.8μm in the centre.

However, the narrow diameter of vessels in capillary beds such as the lungs or peripheral tissues ($\approx$4μm) forces them to adopt a 'parachute' conformation and travel in single-file, bringing much of their surface in close apposition to the endothelium to reduce the distance for diffusional O_2/CO_2 exchange.

A single erythrocyte will undergo conformational changes hundreds of thousands of times during its short 120-day life span.

Erythrocyte Senescence and Removal

Traversing the circulatory system with multiple repeated conformational changes exposes erythrocytes to greater physical insult than many other cell types in the body. Combined with an inability to repair any damage they receive means that any cytoskeletal or membrane damage accumulates rapidly, increasing the risk of membrane rupture.

Ageing or damaged erythrocytes therefore have to be removed from the circulation before they risk rupturing and spilling their contents, explaining their brief 120-day lifespan in the circulation.

The biological system which removes senescent erythrocytes from the circulation is termed the *reticuloendothelial system* (RES) and includes specialised cells resident in the spleen and the liver as well as in other lymphoid tissues throughout the body.

In addition to the RES, the liver, kidneys and some other tissues release a protein into the blood called *haptoglobin*. Circulating haptoglobin is able to bind 700mg/L – 1.5g/L of free hæmoglobin in the plasma to inhibit its damaging effects on tissues and to allow uptake by the RES[80]. The haptoglobin-hæmoglobin complex is also unable to pass through the glomerulus, thus preventing renal tubular occlusion by deposition of free hæmoglobin.

The functioning of the RES and plasma haptoglobin maintains the plasma free hæmoglobin level below 100mg/L - less than 0.1% of the amount of hæmoglobin concentration carried within erythrocytes – a ratio termed $\left(\frac{\Delta \mathrm{Hb}}{\mathrm{Hb}}\right)$[80].

The Impact of Hæmolysis

Should erythrocytes lyse *en masse* so as to saturate these natural scavenging mechanisms, free hæmoglobin is highly oxidative, osmotically active, and also able scavenge free nitric oxide[80, 81] – a key vasodilator. Hæmolysis also results in the showering of multiple embolic lipid membrane fragments into the circulation which can occlude small vessels together with precipitated hæmoglobin.

Thus the result of hæmolysis is a combination of osmotic, oxidative, vasoconstrictive, embolic and anoxic damage to end organs, particularly organs with poor collateral circulation such as the kidneys and lungs. Patients suffering hæmolysis experience a range of symptoms including hæmoglobinuria (the presence of hæmoglobin in the urine), abdominal pains, elevated blood pressure and jaundice[80–83].

3.2.2 The Clotting Cascade

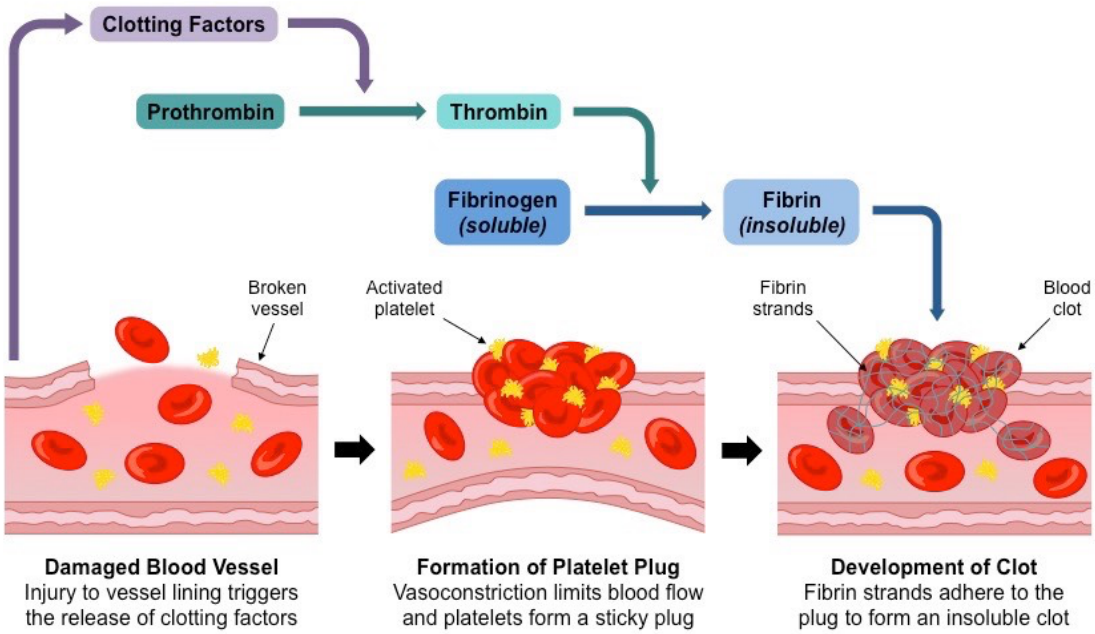


Figure 3.2: A simplified overview of the process of blood coagulation cascade demonstrating the production of a thrombus (clot) from cellular and acellular elements of the blood

Organisms with circulatory systems have evolved a system of blood coagulation as an adaptation to stem the flow from leaks, and to repair any damage to vessels resulting from trauma.

The human blood clotting system is highly complex and involves multiple proteins and co-factors that act as promotors and inhibitors in order to maintain the balance of the

system on a knife-edge, waiting to be either rapidly deployed or rapidly reversed. The majority of the key proteins are synthesised in the liver and comprise a large proportion of the acellular proteinaceous components of the blood[1], along with immunoglobulins, albumin and others[2].

The clotting cascade is triggered by two major routes, termed the *tissue factor pathway* and the *surface pathway*[3]. Both of these pathways converge to form a fibrinous mesh that seals leaks and entraps platelets, erythrocytes and leukocytes to initiate repair.

The *surface pathway*, the less active of the two, is triggered by activating proteins being exposed to repeating charged molecules. In a biological system this is sub-endothelial collagen. Damage to the endothelium exposes the collagen and therefore triggers this pathway.

The *tissue factor pathway* is triggered by the presence of a specific clotting initiator called *tissue factor* on leukocytes and subendothelial smooth muscle cells. The presence of this initiator is a more potent activator of the clotting cascade than collagen.

The Role of Platelets in Clotting

Platelets are the major cellular component involved in the clotting cascade. They derive from large nucleated precursor cells residing within the bone marrow termed 'megakaryocytes' and are released by budding off from the surface. They are anucleate and packed with large numbers of pre-formed granules which contain a highly thrombogenic 'payload'.

Platelets play an important role in the clotting cascade. If activated by adhesion to subendothelial structures, through inflammatory cytokines, the presence of *tissue factor*[4], or high shear stresses[84], platelets are triggered to release their granules. The chemicals within these granules cause:

- other platelets to degranulate and aggregate
- the formation of a fibrin mesh through the clotting factor cascades
- leukocyte migration into the region to begin the process of repair.

The end result of the combination of the fibrinous mesh and entrapped platelets is termed a *thrombus*.

[1] Fibrinogen, the immediate precursor to the fibrin that forms the mesh-like 'clot', comprises some 7%–10% of the total protein content of blood

[2] Blood without its cellular component is called *plasma*, and if plasma is artificially coagulated so as to remove the clotting proteins, it becomes *serum*

[3] Also known as the *extrinsic pathway* and the *intrinsic pathway*, respectively

[4] It was previously believed that platelets generated tissue factor, but this now appears not to be the case

The Impact of Thromboembolisation

Emboli are abnormal objects travelling within the blood stream. They can be anything from air bubbles to fat, to bone fragments or even components of artificial heart valves.

Small particles that break free from a thrombus, or small thrombi that form in the circulation, are capable of becoming lodged in vessels of lesser diameter than themselves. If the body's natural thrombolytic or cellular mechanisms fail to consume the embolus in time, this occlusion can lead to downstream ischæmia and tissue death.

This is particularly serious in organs with poorly collateralised vasculature such as the *vasa rectæ* of the kidney (page 60), explaining why microembolisation can be particularly harmful to renal function.

3.2.3 Shear Stresses and Turbulent Flows

Early animal experiments using artificially induced vessel turbulence[85] demonstrated a causative relationship between turbulent flow and arterial thrombosis. Similarly, the exposure of erythrocytes to sudden changes in membrane stresses, such as those within atherosclerotic vessels or peristaltic pumps, results in erythrocyte damage and hæmolysis[81].

However, the cellular and acellular mechanisms underlying these phenomenon have taken longer to come to light, and are still subject to great scrutiny owing to the burden of arterial and cardiac disease in the population.

We now understand that hæmolysis and thromboembolisation share a common ætiology - the presence of abnormal flows and high shear stresses[82, 86–88].

Two important characteristics describe the pathogenesis of blood flows – the *shear rate* and the *wall shear stress* of the flow.

Consider that a fully developed laminar Poiseuille flow in a channel adopts a parabolic flow profile. The *shear rate* describes the relative velocities between the 'shells' of flow and describes the cell velocity in free flow, and the *wall shear stress* describes the transverse forces acting between the stationary vessel wall and the liquid flowing past it. Shear stress (measured in Pascals or Dynes/cm^2) is the product of the shear rate (s^{-1}) and the dynamic viscosity of the liquid (μ).

Physiological shear rates in large vessels lying in the range of 20 – 2000 s^{-1} have no effect on platelet activation and thrombosis or hæmolysis.

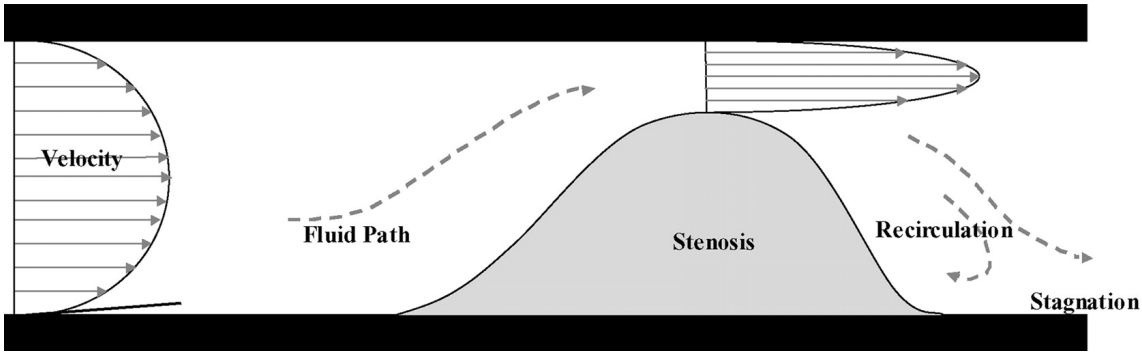


Figure 3.3: Diagram of blood flow through a constriction point demonstrating the increase in shear rates through constrictions and post-stenotic turbulence. From[84]

However, rapidly narrowing vessels such as those resulting from atherosclerosis, have greatly elevated velocities and shear rates (5,000 – 20,000 s^{-1}), and therefore wall shear stresses through the constriction point. This promotes platelet binding through the up-regulation of specific surface receptors[86]. Furthermore, erythrocytes subjected to these high shear stresses release purines (ATP and ADP) which are potent stimulators of platelet activation.

In the case of hæmolysis, we know that erythrocytes are naturally able to reversibly deform in response to changes in flow rates and vessel diameter (page 78). However, high magnitude deformations resulting from high shear stresses exceed their ability to recover and leads to membrane damage and lysis.

Various studies have tried to identify the shear stress threshold for hæmolysis in laminar and turbulent flows[87], resulting in values between 1,500 – 3,000 dynes/cm^2 (150 – 300 Pa), versus normal arterial wall shear stresses which fall in the region of 70 dynes/cm^2.

It appears therefore, that the threshold for hæmolysis is not a single fixed value. This is partly due to the population effect of erythrocytes in various states of senescence (exemplified by the finding that transfusions with red cells kept for >14 days result in much higher levels of hæmolysis than fresh blood[81]), but also the summative nature of the exposure to shear stresses over time, which will be different for each individual erythrocyte.

Giersiepen *et al.* created a mathematical description of the relationship between the proportion of free hæmoglobin resulting from hæmolysis ($\frac{\Delta Hb}{Hb}$), the shear stress (τ) and the duration of exposure of erythrocytes to this stress (t_{exp}) whilst traversing heart valve prostheses in a mock circulatory loop[89]. He determined that the duration of exposure to shear stress plays a role in the development of hæmolysis, although in a sub-linear fashion compared to the absolute magnitude of the stress.

$$\frac{\Delta \mathrm{Hb}}{\mathrm{Hb}}(\%) = 3.62 \times 10^{-5} \cdot \mathrm{t}_{\mathrm{exp}}^{0.785} \cdot \tau^{2.416} \tag{3.1}$$

Others have since tried to improve upon this estimate with various computational methods (see [90] for an example) but there is still no robust way of predicting hæmolysis rates through simulation.

3.2.4 Summary of Desirable Characteristics of Low-Hæmolysis Blood Pumps

A pump designer must be mindful of the characteristics of blood flow that leads to low rates of hæmolysis and thromboembolisation, and try to mimic these biological mechanisms, for example by:

- Using biocompatible coatings to resist platelet aggregation and/or clotting factor activation
- Removing all components of the clotting cascade from the perfusate (see page 122 for details)
- Using an impeller design that minimises:
  - The wall shear stress of the impeller
  - The dwell-time of erythrocytes in the impeller

Another approach could be to accept a degree of hæmolysis and instead design systems to mimic the reticuloendothelial-haptoglobin system by infusing haptoglobins and removing red cells and degraded erythrocyte membranes before they can damage tissues or form microemboli, while replacing the lost red cells with fresh ones.

Such systems may be possible one day, based upon modified dialysis-type cartridges to act as filters, but this may be unnecessary if the problems can be solved at the pump stage.

We will now consider how different types of blood pump have been developed from a long history of pumping technology.

3.3 Evolution of Pumping Technology

The history of moving liquids dates back to the time of the earliest pre-technological farmers, who created channels for irrigating their crops. As the rivers dropped with the seasons, they undoubtedly struggled with simple buckets and manual labour, desiring a simpler mechanism to relieve them of their labour.

Around 4,000 – 5,000 years ago the Ancient Egyptians began to use a simple bucket on a counterweighted lever known as a '*shadoof*' to raise water from low rivers or wells[91–93]. This was the first practical human 'liquid moving' technology, the precursor to pumping.

The next great step came in around 250BC when Ctesibius (285 – 222BC) invented the reciprocating suction pump in Alexandria, Egypt[1]. His device was based on two closed pistons and a simple valving mechanism in order to drive water along a pipe. It was this mechanism that was used extensively by ancient Greek and Roman firefighters until the knowledge was apparently lost in the burning of the library of Alexandria. By the middle ages, Europeans were back to using human relay lines of buckets and wet cloths to fight fires, until the reciprocating pump mechanism was 're-discovered' in the 1700s.

Archimedes (287 – 212BC) was a contemporary of Ctesibius, although he lived on the other side of Mediterranean in the Greek city of Syracuse, Sicily. There is no doubt as to his contributions to mathematics and physics in general, but his contribution to pumping technology with the invention of the 'Archimedes screw' is probably apocryphal. The mechanism of using an enclosed rotating screw lying at an angle to continually draw water from a source was probably in use by the Assyrians and Babylonians some 300-400 years prior to Archimedes' 'discovery' of the mechanism. Indeed, his 'invention' was probably just its introduction from the unknown world of the Levant to the known world of educated *Magna Græcia*.

The next stages of development followed in rapid succession over a few hundred years from the late Renaissance era until the Industrial Age. It is perhaps better to address the developments from this stage onwards according to the type of pumping mechanism employed.

3.3.1 Types of Pump

The two major divisions of pump mechanism are the *positive displacement* type and the *dynamic* (or *centrifugal*) type which developed many centuries later.

[1] He also realised air was a substance that behaved like a fluid, invented a water-based musical organ and created cam-operated musical devices and automata for public entertainment. Sadly all of his writings have been lost to history, and he died in poverty

3.3.2 Positive Displacement Pumps

Positive displacement pumps use physical means to impart energy to a liquid, often involving the alteration of the volume of a chamber into which the liquid is drawn and from where it is then expelled. These pumps are often able to operate against large pressure heads as they pump a constant volume of liquid per action, and the driving mechanisms can be adjusted to alter the rate or pressure head as necessary.

Furthermore, positive displacement pumps can also be used for highly viscous and even non-liquid applications such as moving powders or industrial aggregates.

There are multiple different families of positive displacement pumps using the following mechanisms:

- Reciprocating linear mechanisms – pistons, diaphragms
- Rotary blades – screws, interlocking gears, vanes and peristaltic devices

Reciprocating Mechanisms – Piston Pumps

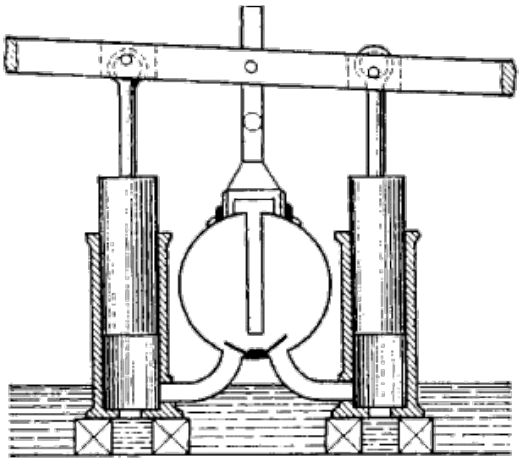


Figure 3.4: A simple dual-piston pump

After the Archimedes Screw, this design has longest history of any pump design. It was at its peak the most widely deployed pump technology in the world, enabling mining and the machines that powered the industrial revolution of the 1800s.

The principle is straightforward and remains as invented by Ctesibius - a piston drawn along a cylinder sucks in liquid through a valve, and pressing down on the piston then expels the fluid through the same or a different valve. Ctesibius' knowledge wasn't completely lost, as evidenced by illustrations of a dual-piston mechanism in a 1475 treatise on civil and military architecture by Italian Renaissance engineer Francesco di Giorgio Martini.

Ctesibius' design was re-designed with modern improvements by Otto von Guericke (1602 – 1686) in 1650 and Sir Samuel Morland (1625 – 1695) in 1675.

Thereafter followed a succession of improvements in the mechanisms for powering such pumps - from manual labour to water mills to steam turbines. Frenchman Denis Papin (c1647 – c1712) is credited with creating the first practical steam powered piston in 1690[1] and later published *The New Art of Pumping Water by using Steam* in 1707, following his work on using steam engines to pump water into a canal and to supply water fountains at the Landgraves Palace of Marburg[2].

Unfortunately, Papin died in poverty on an unknown date with his inventions having gone nowhere in his lifetime. Soon after Papin's death however, Thomas Newcomen (1664 – 1729) put the principles of his steam pump to great use in draining Cornish tin mines of water, and by 1733 around 125 of the devices had been installed across industrial Britain and Europe. By 1775 some 600 Newcomen engines had been built.

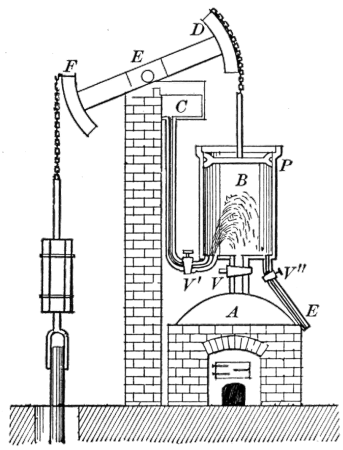


Figure 3.5: Newcomen's Steam Pump [94]

The most famous inventor of the steam era is undoubtedly Sir James Watt (1726 – 1819) and his business associate Matthew Boulton (1728 – 1809). Watt was repairing a model of Newcomen's engine at the University of Glasgow when he struck upon his idea of condensing the steam in a secondary cylinder attached to the boiler rather than the boiler itself. His first successful experiment in 1765 led to a full-sized working engine in 1775 which required only a quarter of the coal of Newcomen's design.

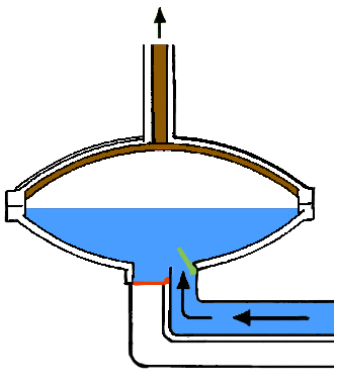
As an aside, in 1781 Watt invented an improved mechanism which was capable of converting the reciprocating motion of a steam-driven piston into rotary motion to drive a vaned rotary blade within a closed housing, thus creating a steam-driven rotary pump, rather than the previous linear piston-driven devices of Newcomen and others. This new rotary motion also found use in running textile mills, boring holes and driving bellows to ærate blast furnaces across Britain, cementing Watt's place in history.

Modern piston pumps also include precision syringe drivers used in medicine for volumetric drug delivery by a micro-screw mechanism.

[1] Although the invention of the first practical steam engine is credited to Thomas Savery (c1650 – 1715) in 1698

[2] Papin is also credited with the invention of the steam cooker

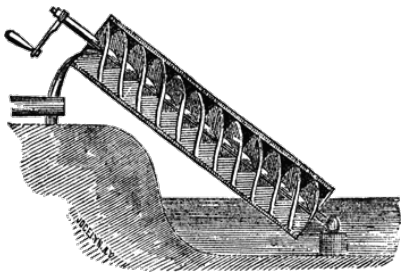
Reciprocating Mechanisms – Diaphragm Pumps



Interestingly, the development of the first diaphragm (or membrane) pump came late in the history of pump technologies, following the work of Jacob Edson (1786 – 1870) in 1859. Edson was granted US Patent US26025, and his device was first used for marine bilge pumping[1].

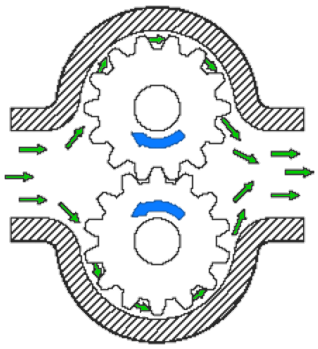
The diaphragm pump mechanism is essentially a version of the piston pump. It relies on a single chamber with inlet and outlet valves. By drawing on the diaphragm, the volume of the chamber increases, opening the inlet valve and drawing in the liquid. Force is then applied to the diaphragm which closes the inlet valve and opens the outlet valve, expelling the liquid from the chamber. These are highly efficient and minimal designs with few moving parts, and are beginning to find uses in microscopic, MEMS devices.

Rotary Blades – Screw Pumps



The best and most lasting example of this type is the Archimedes Screw, discussed in the introductory section of this Chapter on page 85. This type of pump still finds use in regions requiring the transfer of water from lower-lying sources into irrigation ditches, and interestingly in the Netherlands for continuous drainage of polders.

Rotary Blades – Gear Pumps

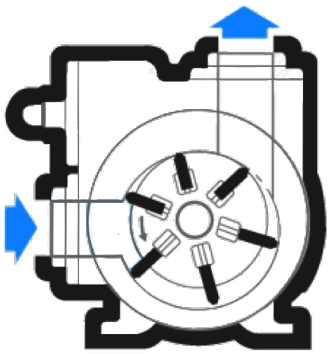


Nicolas Grollier de Serviére (1596 – 1689) is credited with the first design for a gear pump employing two interlocking toothed gears to impart energy to water. He described the ability to turn the gears by the motion of an external crank and that this device could be used in both a 'sucking' and 'forcing' configuration([91] pp. 285).

The mechanism and its many derivatives remain popular to this day, particularly in hydraulic machinery.

[1]Edson went on to found the highly successful Edson Marine Company which still exists today

Rotary Blades – Rotating Vane Pumps



Agostino Ramelli (1531 – c1610) recorded over 100 designs of water pumps and water well mechanisms, and is credited in particular with the invention of the rotating vane pump in 1588, based on a mechanism much like a water wheel contained within a fixed housing.

This mechanism was essentially re-invented in 1876 by Charles Barnes (1813 – 1884), and then improved in 1899 by Robert Blackmer (c1862 – 1957) who developed the sliding vane design shown here, whereby the tips of the spring-loaded vanes on the eccentric rotor always remain in contact with the housing, even after significant wear. The vanes of these pumps sweep the periphery of the housing, drawing fluid from the inlet and driving it towards the outlet. Blackmer's company (incorporated 1903) still produces these pumps to this day.

Rotary Blades – Peristaltic Pumps

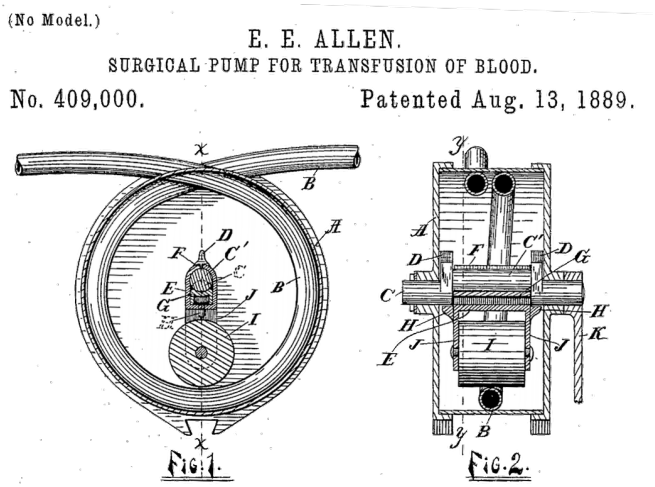


Figure 3.6: Eugene Allen's 1881 Peristaltic Pump Design

The concept of the peristaltic pump is now just over 150 years old, and derives from the rotating vane pump. It was first patented by Rufus Porter (1792 – 1884) & J.D. Bradley in 1855 (US12753) for pumping the bilge from ships, or viscous liquids such as molasses. A later single-roller design was patented in 1881 by Eugene Allen (US409000) with the specific purpose of blood transfusion in mind (the image above shows one of his later designs). This pump type also finds industrial use in handling hazardous chemicals,

biological materials and solid slurries because the driving rotor never has to come in contact with the substance being pumped.

In a peristaltic pump, the spring-loaded vane is in the form of a rotary roller which compresses a plastic tube against the wall of the housing. As the vanes rotate, aliquots of fluid are drawn from the inlet source and expelled through the outlet.

3.3.3 Dynamic Pumps

This second major category of pumping technology relies on a moving fan or impeller immersed in the liquid itself to impart the energy to drive the liquid.

Dynamic pumps require more careful matching of the designs of the impeller and the housing so as to optimise this energy transfer in order to create pressure and/or flow at the outlet, involving a degree of precision engineering which was unavailable until the 19th century. Likewise, the development of the dynamic pump came much later than that of positive displacement pumps as the primary motive source had to be rotational and capable of maintaining high angular velocities.

One of the key properties of modern dynamic pumps is their versatility. They can be customised to occupy nearly any niche according to the type of liquid to be pumped and the pressure and flow qualities desired, with the caveat that in general dynamic pumps are not able to generate pressure heads as great as those of positive displacement pumps, and their flow rate and pressure head can be adversely affected by increased fluid viscosity. This makes positive displacement pumps still better suited for metering, non-contact and high pressure applications.

There are two main configurations of this type dynamic pump – the centrifugal and axial flow type, describing the direction of liquid flow in relation to the impeller.

Dynamic Pumps – Centrifugal Type

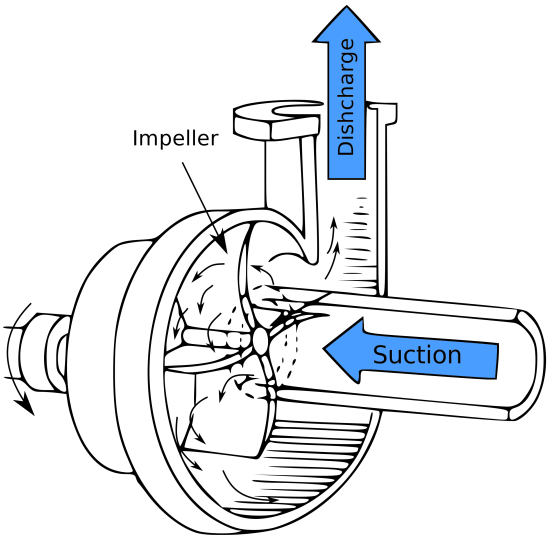


Figure 3.7: Basic design of a centrifugal pump

In the centrifugal configuration (hence the name *centrifugal pump*), the fluid is drawn in along the axis of the impeller, from which it gathers kinetic energy and is then expelled radially into the diffuser (depending on the design) and then into the volute, where this kinetic energy is converted into both forward flow and a pressure head. This is the inverse of the behaviour of turbines, which developed from the same design principles.

Denis Papin is credited with not only creating the first practical steam powered piston (page 87), but also the first centrifugal pump for ventilating coalmines in 1687. Unfortunately, much like his other work Papin saw no uptake of his designs despite a successful demonstration to the King of England on the banks of the Thames, but we can see the first principles of a spiral housing and a central bladed impeller.

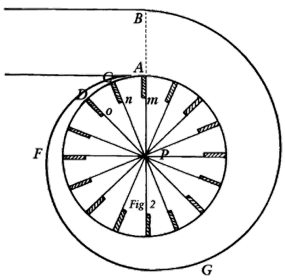


Figure 3.8: Papin's *c*1705 design

It wasn't for more than another 100 years before Papin's principles were reapplied successfully in the form of the 1818 'Massachusetts Pump', a (very inefficient) straight-bladed centrifugal pump under the power of steam. The name of the inventor of this pump has unfortunately been lost to time. Papin's work was again rediscovered by W.D. Andrews who was granted a patent in 1839 for a centrifugal pump design with tapered vanes and a spiral volute.

The next stage of historical development included the application of steam power to centrifugal pumping by James Gwynne (1804 – 1850) in 1846, working from W.D.

Andrew's US patent. Gwynne's company survives to this day as a major pump manufacturer.

Around 10 years before this, in 1827, Benoît Fourneyron (1802 – 1867) developed a major technological improvement in water turbine technology when he used horizontal water wheels with recurving blades to generate 6 horsepower, greatly outperforming the traditional paddle-type water wheels of the time. The concept of the curving blade was applied to the centrifugal pump by John Appold (1800 – 1864) in 1851.

Thereafter, Arthur Giesler patented the multistage centrifugal pump in 1905 (US948292), designed to achieve higher compression and pressure heads than single stage devices.

From this point until the modern day, various improvements to centrifugal pumps have responded to the introduction of electric power, materials technology and computer modelling to result in the high capability industrial pumps now occupying a vast array of roles across the world.

Dynamic Pumps – Axial Type

The most common type of axial impeller is the propeller mounted on millions of ships worldwide. Here the liquid is drawn in from the forward face of the impeller and is accelerated along the same axis with minimal radial excursion before being propelled out the rear, thus either driving the ship forward in the water, or acting to pump the liquid.

The first effective axial pump was the Archimedes Screw, but this device also sits comfortably in the category of a rotary bladed positive displacement pump owing to its fixed shroud and close apposition of the screw against the walls of the pump.

The first ever true ship propellers, improving on the side-mounted paddles of the paddle-steamers of the day, were in fact Archimedes-type screws. These were being trialled on ships in the early 1800s until a serendipitous accident befell Francis Pettit Smith (1808 – 1874). Smiths screw-driven boat was being tested on the Paddington Canal when one day in 1837 his wooden screw consisting of two full turns fractured in half and his craft actually *accelerated* on the single-turn screw.

Smith is also famous for a demonstration later that year when his single-screw craft 'ran circles around' a Royal Navy flotilla steaming along in stormy seas off the North Kent coast. Just 8 years later Isambard Kingdon Brunel's *SS Great Britain* crossed the Atlantic on screw drive.

Propellers then took centre stage during the quest for flight in the late 1800s and early 1900s. Alongside multiple practical designs came the mathematics and theories to describe

propellers and aerofoils in general and designs stabilised throughout the latter part of the 20th century.

Axial turbopumps also found use in the rocket programmes of NASA and the Soviet Union through the 1950s onwards.

Then in the late 1980s and early 1990s, the perfect combination of improvements in materials, miniaturisation and computational fluid dynamics brought axial flow pumps to the centre of attention for medical blood pumping and led to the successful work of Dr. Michael E. DeBakey, Dr. George P. Noon and a team of NASA engineers in pioneering the second generation of *Left Ventricular Assist Devices* (LVADs). This will be discussed further on page 102.

3.4 Specialisation into Blood Pumps

From the long history and vast array of pumps designed for industry, only a few subtypes have ever translated over into being medically useful. Indeed, in many cases it appears to have been an afterthought that a certain industrial pumping mechanism might find use in the medical field.

The development of medical blood pumps have tended to follow specific historical clinical needs. In the late 1800s, the primary driving force for medical pump innovation was the need to simplify and augment the practice of blood transfusion.

But from the late 1930s onwards the primary motivation to create smaller, better blood pumps has been a response to:

(i) Cardiothoracic surgery – for providing a totally artificial extracorporeal circulation to allow surgeons to operate on the heart or lungs over the course of 4 – 12 hours[1]
(ii) Managing heart failure – where LVADs can provide a new lease of life for certain patients, and act as a bridge for transplantation for others

The focus of medical pump technology may shift further from this point to the long-term extracorporeal support of individual organs for transplantation, or for bioreactors growing new organs from autologous stem cell populations. Only time will tell.

Here we will explore those types of pumping device which have found a successful role for pumping blood for clinical use. We will not cover the gravity devices or early reciprocating piston pumps that were tested in animal experiments of the 1800s, nor the various historical experiments of perfusing one animal from the heart and lungs of another.

[1]Cardiothoracic procedures beyond 12 hours in duration are rare, and are associated with increased rates of morbidity and mortality

3.4.1 Positive Displacement Pumps

Two of the subtypes of positive displacement pump have found use in blood pumping. These are the roller, or *peristaltic* pump, and the diaphragm pump. Each type now occupies its own niche in clinical medicine.

Roller (Peristaltic) Pumps

Figure 3.9: A bank of peristaltic pumps on the Sorin S5™ cardiopulmonary bypass machine, LivaNova PLC, Milan, Italy

The first explicit example of a pumping technology being used for moving blood came with Eugene Allen's 1881 patent for a hand-cranked blood transfusion pump (US409000 – see page 89 for details). This was a peristaltic-type pump based on a modification of Porter and Bradley's earlier industrial design of 1855. Transfusion in the late 1800s previously involved the donor lying beside the recipient and their blood being directly transferred using tubing or syringe-type devices.

This new pump meant the doctor no longer had to risk coming into contact with the blood being transfused, and greatly sped up and simplified the process. Transfusion reactions (including hæmolysis) were known and common at the time, but remained inexplicable until 1901 when Karl Landsteiner (1868 – 1943) discovered the concept of 'blood groups' in humans[1].

[1] It could be argued that this work planted the seeds of the entire field of immunology and the recognition of 'self' and 'non-self', a topic covered earlier on page 25

In 1899, Charles Truax (1887 – 1935) licensed Allen's patent and modified his designs to incorporate two rollers to generate a smoother flow, as well as adding extra attachments for other medical uses[95].

Amongst the many others to improve upon Allen's design was Michael E. DeBakey (1908 – 2008), who came across the pump as a medical student in 1932 and improved upon it by preventing the rubber hosing from progressing around the device as the crank was turned.

DeBakey's name became forever synonymous with blood pumps in the late 1930s when his modified roller pump was taken up by the pioneer and inventor of cardiopulmonary bypass surgery, John Heysham Gibbon, Jr. (1903 – 1973).

Gibbon performed the first successful cardiac operation on total cardiopulmonary bypass in 1953 with an oxygenator and perfusion circuit designed by engineers from IBM[95] and incorporating DeBakey's peristaltic pumps.

Were they to step into a modern cardiothoracic operating theatre, Gibbon and DeBakey would find that little has changed since their day. A battery of multi-roller peristaltic pumps form the basis of most modern bypass machines to control various flows to and from the patient. These include pumps for the systemic circulation, blood scavenging and the delivery of specialised solutions to keep the heart paralysed, as well as backup pumps in the event of failure.

To replace the systemic circulation, these pumps must be capable of meeting the 7L/min flow requirements at up to 500mmHg pressures[77], and have manual controls to enable battery operation in the case of power cuts – a rare but potentially life-threatening problem in operating theatres.

Fortunately, these qualities are within the capabilities of even small peristaltic pumps.

Advantages & Disadvantages of Peristaltic Pumps

The advantages of peristaltic pumps include:

- No dependence on pre-load or afterload, meaning that the pump can eject into nearly any pressure head and self-fill from zero pressure reservoirs
- The occlusive nature of the roller mechanism prevents any backflow through the device
- Metering of flow can be adjusted to a fine degree owing to the direct coupling between the rate of roller rotation and blood aliquot delivery

However, these must be balanced against certain problems which are specific to this pump design[77]:

- The degree of compression of the rollers against the tubing must be carefully adjusted and monitored to balance the need to generate pressure and forward flow against the risks of traumatic hæmolysis and frictional spallation (release of microparticles from the silicon or PVC tubing, an effect worsened by hypothermia).
- The risk causing microscopic stress fractures in the plastic tubing, introducing air emboli into the blood flow
- High negative pump pressures generating cavitation effects which include all the risks listed above in addition to severe hæmolysis

In the field of organ transplantation, peristaltic pumps are also the pump of choice for the LifePort® Kidney Transporter[1] which uses a hypothermic cell-free crystalloid perfusate.

Diaphragm Pumps

Following Jacob Edson's original filing for a diaphragm-based bilge pump in 1859 (US26025), there were a number of sporadic patents describing the possible use of this technology for pumping blood through the late 1950s to the 1960s[2].

The introduction of the *Intra-Aortic Balloon Pump* (IABP) for cardiac surgery in the early 1960s could reasonably be claimed as the first use of a diaphragm pump in medicine, with the nitrogen-inflated balloon acting as the diaphragm, and the walls of the native aorta acting as the pump housing.

However, the first diaphragm pump that would be recognised by its inventor would have been the CO_2-driven external LVAD designed and implanted into a number of patients between 1963 and 1966 by Dr. Michael E. DeBakey[96].

An LVAD acts to supplement the function of the failing left ventricle (LV), and works in parallel to it by drawing blood from the LV and ejecting it into the aorta, either synchronously or asynchronously with the native heartbeat. As such, an LVAD does not need to generate as great a flow or pressure head as a bypass pump – more on the order of 3L/min with a head rise of 100mmHg, to supplement the failing native cardiac output.

The most famous of DeBakey's early LVAD patients was a 37 year old woman who required the device for 10 days after mitral and aortic valve replacements before her own heart had recovered sufficiently to allow her to leave hospital. No doubt many of the recipients who weren't included in DeBakey's 1971 case reports had not been so lucky.

Development of these devices continued through the 1970s and 80s, and resulted in the commercially-available PVAD™ (1982) and HeartMate I® (1986) devices[3] as well as the

[1] Organ Recovery Systems, Des Plaines, Illinois, USA
[2] Such as US3007416 (1958) and US3048165 (1959)
[3] Thoratec Corp., Pleasanton, California, USA

now-defunct Novacor N1000[1].

These first-generation devices were all single-ventricle assist positive-displacement devices comprising a sterile 60ml – 80ml blood chamber with inlet and outlet valves, and a pneumatically-driven diaphragm. The sterile part of the device was either implanted in the patient's abdomen or worn externally in bed-ridden patients. They were all driven by pneumatic lines connected to an external compressor. Some of the later first-generation LVADs converted from pneumatic to electrical drives, but still relied on a positive-displacement diaphragm pump methodology. However, this meant patients could now be mobile with a shoulder-carried battery pack.

Dr. Robert Jarvik (*b.*1946) also used air-driven diaphragm pumps for the extreme application of total artificial heart replacement, releasing his Jarvik-7 device in 1982, later renamed the 'CardioWest™ temporary Total Artificial Heart' in 1990. This device now allows up to 80% of carefully selected recipients to live until a transplant becomes available[97].

The electrically-driven AbioCor Total Artificial Heart[2] was released in 2001 and trials showed it was able to extend life up to 17 months in some recipients. Unfortunately, it was only implanted in 15 patients before being withdrawn due to issues of safety and efficacy. The CardioWest™ device remains on the market, although it requires the patient to remain bed-bound and connected to an air compressor.

Advantages & Disadvantages of Diaphragm Pumps

The first-generation LVADs came with a number of problems, not least[98]:

- Pneumatic machinery limiting patient freedom
- Pneumatic line rupture
- Membrane leakage or rupture from repeated strain flexing
- High risk of infection via transcutaneous lines and pipes
- Hæmolysis
- Thrombosis and embolisation - although the HeartMate I® device was at lower risk due to formation of a pseudo-intimal layer

Of course, these risks were overwhelmed by the ability of these devices to sustain life where no others could, doubling the 1-year survival of the recipients when compared to medical therapy alone.

[1] World Heart, Inc., Oakland, California, USA
[2] AbioMed Inc., Danvers, Massachusetts, USA

3.4.2 Dynamic Pumps

Both subtypes of dynamic pump, the centrifugal and axial types, are employed in clinical medicine for cardiothoracic bypass and as second and third-generation LVADs, respectively. As we will see, dynamic pumps do have specific functional advantages as well as disadvantages when compared against positive displacement pumps occupying similar niches.

Centrifugal Pumps

The first centrifugal pump that found use in cardiopulmonary bypass circuit was the Bio-Pump®[1] in 1976, four decades after John Heysham Gibbon, Jr. began using DeBakey's modified peristaltic pump. Centrifugal pumps are now able to take on most pumping capabilities in bypass machines, except for suction where positive-displacement peristaltic pumps still excel.

Centrifugal blood pumps must still be able to achieve the same pressure/flow characteristics of peristaltic pumps – namely flow rates of 7L/min at pressures reaching 500mmHg. In order to achieve this, a variety of designs have emerged over the years, each claiming specific advantages over their commercial competitors. All such pumps fall into the following categorisation scheme:

Bladed Impeller

- Straight Blades
  - Sarns™ [2]
  - Nikkiso HPM-15 [3]
  - MedTech Dispo™ [4]
- Curved Blades
  - St. Jude LifeStream[5]
  - Cobe-Stöckert[6]

Non-Bladed Impeller

- Flat Rotor
  - Capiox® SP[2]
- Stacked Cones
  - BPX-80 Bio-Pump® Plus[7]

The bladed impeller pump types in the list above operate in very much the same way as Denis Papin first imagined back in 1705 (illustrated on page 91), although benefitting from modern materials technology including integration of anticoagulant heparin molecules onto the plastic surfaces, and computational fluid dynamics to optimise the designs.

[1] BioMedicus Inc., now Medtronic Bio-Medicus Inc., Eden Prairie, Minnesota, USA
[2] Terumo Corporation, Shibuya, Tokyo, Japan
[3] Nikkiso Co. Ltd., Tokyo, Japan
[4] MedTecHeart Inc., Tokyo, Japan - A pre-commercial VAD that uses a magnetically levitated impeller
[5] St. Jude Medical, Inc., Cardiac Assist Division, Chelmsford, Massachusetts, USA
[6] Cobe Cardiovascular Inc., Arvada, Colorado, USA
[7] Medtronic Bio-Medicus Inc., Eden Prairie, Minnesota, USA

The more interesting centrifugal pump designs are those in the second column – those employing *non-bladed impellers*.

The Terumo Capiox® SP

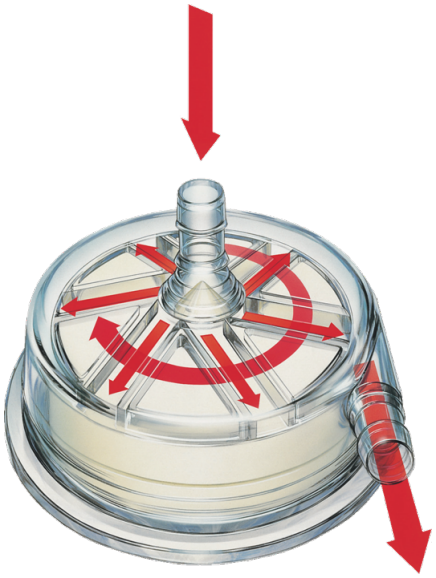


Figure 3.10: The Capiox® SP from Terumo[99]

The Capiox® SP uses a heparin-bonded rotor comprising 6 tunnels of constant area extending radially from the central inlet. Terumo term this a 'straight-path' design and the intention is to reduce areas of stasis whilst also reducing hæmolysis. The design harvests kinetic energy directly from red cells accelerating radially outwards through centrifugal force.

Because the blood does not encounter any immobile surfaces once entrained within the tunnels, it is clear that there will be no abrupt shear boundaries in their path such as with a traditional impeller washing through a volume of blood. The only region of abrupt acceleration occurs at the inlet, where the velocity of the incoming blood is at its lowest.

The Medtronic BPX-80 Bio-Pump® Plus

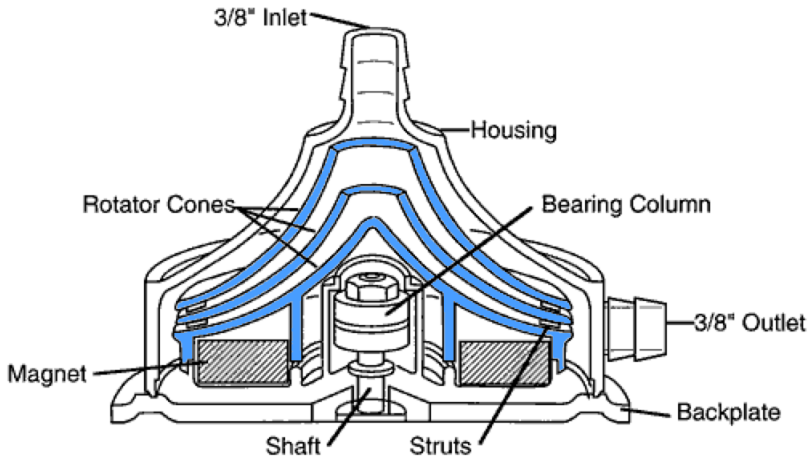


Figure 3.11: The BPX-80 Bio-Pump® Plus from Medtronic

The BPX-80 Bio-Pump® Plus is the second of the non-bladed types, and coincidentally the oldest of the centrifugal blood pumps, taking its design leads from the original BioMedicus Bio-Pump® introduced in 1976. The impeller comprises a series of 3 stacked cones rotating within a conical housing.

In this design, kinetic energy is imparted to the blood through a combination of viscous drag and gravity. As the discs rotate, boundary effects draw the blood in the direction of rotation. The longer the duration of contact and the higher the angular velocity of the cones, the more kinetic energy will be imparted to the blood. In this way, the acceleration can be continuous and smooth, with constant shearing forces applied so as to minimise hæmolysis. The effect of shear stress on hæmolysis was covered in more detail on page 76.

The flow profile that develops between the uppermost cone and the stationary inner aspect of the conical housing can be described as a Couette flow. Between the other two co-rotating inter-cone spaces, the flow will adopt a profile dependent on the angular velocity of the discs and the kinematic viscosity of the blood – a phenomenon investigated at length by Breiter & Pohlhausen in 1962 and Hasinger & Kehrt in 1963. This is covered in more detail from page 115 onwards.

As a result of these different flow profiles, one study has shown that shear rates between the co-rotating inner cones of the BPX-80 Bio-Pump® Plus are largely constrained to between 100 – 400 s^{-1} at flow rates of 400 – 2000ml/min, compared to the rotational-speed-dependent shear rate between the outer cone and the stationary housing of approximately 2.57 s^{-1}/rpm, or $\approx$ 5,200 s^{-1} at peak flow – 13$\times$ greater than that between the co-rotating cones[77], and within the range of shear rates found in pathological blood vessels[88].

This is likely to explain the findings of Kawahito *et al.* [100] who performed a head-to-

head comparison between the previous incarnation of the Bio-Pump® and the then-new Capiox® pump in 1997, and found that the Bio-Pump® resulted in 3× more hæmolysis than the Capiox design.

It is interesting to note that this centrifugal pump is also finding use as a short-term extracorporeal VAD for those patients who cannot immediately wean from bypass support[101].

Advantages & Disadvantages of Centrifugal Pumps

The advantages of centrifugal pumps include[77]:

- The dependence on preload and afterload means that the pump cannot overpressurise positively or negatively, leading to less risk of cavitation
- The non-contact nature of the pump mechanism means that there is no spallation of fine plastic particles into the circulation
- The mechanism of blood acceleration also leads to less hæmolysis than peristaltic pumps
- Microemboli may collect in the device and fail to be ejected to the circulation, although this is not guaranteed and is not a 'safety feature'

However, there are also problems which are specific to centrifugal pumps:

- Magnetically-coupled impellers may decouple from the desired angular velocity if they are required to move faster or slower than their design range
- Flow through the pump is strongly dependent on afterload, such that centrifugal pumps are less able to generate large pressure heads than positive-displacement pumps
- Centrifugal pumps are non-occlusive (non-valved) such that it is possible to get backflow through the device. This is more commonly seen shortly after the pump is turned off

In the field of organ transplantation, centrifugal pumps are the pump of choice for the OrganOx metra™ normothermic liver perfusion device[1] which uses oxygenated blood as a perfusate.

[1] OrganOx Ltd., Oxford, UK

Axial Pumps

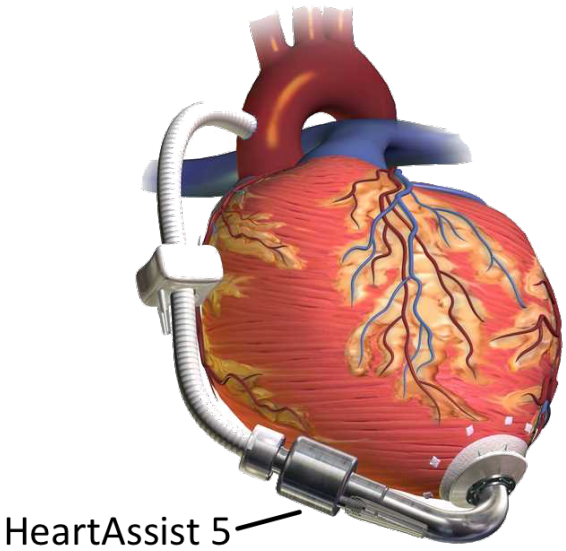


Figure 3.12: The HeartAssist 5™ from ReliantHeart Inc. This 92g, 7cm×3cm device draws blood from the apex of the failing left ventricle and pumps it directly into the aorta

Continuous axial flow pumps have almost completely replaced the first generation of pulsatile VADs[98] following the first implantation of the DeBakey-Noon VAD in 1998. This device was the result of a decade-long collaboration between Dr. Michael E. DeBakey, his colleague at Baylor College of Medicine Dr. George P. Noon, and a team of NASA engineers. This work was then continued under the auspices of a commercial company[1], resulting in the current HeartAssist5® pump, shown in the image above.

This new design heralded a sea-change in VAD technology. No longer were patients drowned out by loud pumps, chained to beds by lines attached to bulky air compressors, or exposed to the risks of percutaneous infection. Instead these second-generation VADs[2] are fully-implantable in a wide range of recipients, irrespective of body size. Patients can now leave the hospital and go about their daily business connected only to a shoulder-carried battery pack. Continuous axial flow VADs are also a licensed 'destination therapy' for a large proportion of heart failure patients who will never receive a transplant[98].

The major axial VADs on the market include the:

- HeartMate II [3] (20,000 implanted worldwide)
- Jarvik 2000® [4] (900 implanted worldwide)

[1]ReliantHeart Inc., Houston, Texas, USA
[2]There are also 'third-generation' VADs emerging, which use magnetolevitation of their impellers and come in both axial and centrifugal configurations.
[3]Thoratec Corporation, a subsidiary of St. Jude Medical, Inc., Chelmsford, Massachusetts, USA
[4]Jarvik Heart, Inc., Manhattan, New York, USA

- HeartAssist5® [1] (≈100 implanted worldwide)
- CircuLite Synergy[2] (≈100 implanted worldwide)
- Procyrion Aortix™ [3] – a very exciting 6mm-diameter catheter-deployed VAD undergoing pre-clinical testing

The mechanism underlying all these devices is the axial flow turbine, essentially a modified propeller which has borrowed heavily from advanced ærospace engineering and materials technology. At its core is a magnetically or hydrodynamically levitated rotor which spins at rates of 8,000–15,000 rpm within a fixed chamber.

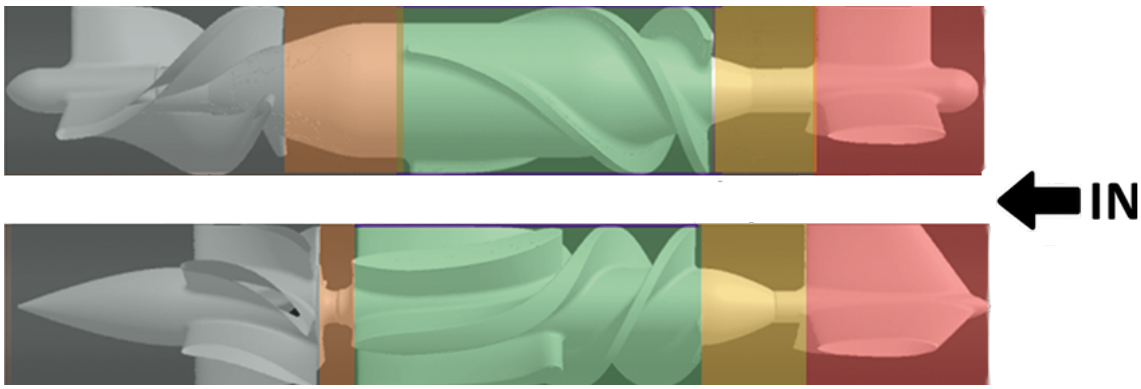


Figure 3.13: Turbines from two commercial VADs. Top: HeartMate II. Bottom: HeartAssist5®

In an axial pump, blood is first drawn into the rotor past a stationary flow straightener (red), and is then fed into an impeller (green) designed to impart kinetic energy to the blood. The blood finally passes a stationary diffuser (grey) which is designed to convert the gain in kinetic energy into pressure and straighten the flow into the outlet. The orange/yellow parts are the impeller's stators.

The use of a high speed turbine for blood pumping would seemingly appear to increase the rate of hæmolysis in the system, yet counterintuitively, these devices appear to cause very low rates of hæmolysis, and function effectively as long-term VADs.

The reasons for this may lie in equation 3.1 by Giersiepen *et al.* from his 1990 paper[89], discussed in greater detail on page 83. In essence, although the turbines rotate at very high speed and with large shear stresses, the exposure time of erythrocytes to these stresses is only a fraction of the time that they would reside within a traditional centrifugal pump, thus reducing the integral shear stress upon each individual erythrocyte.

The second concern held by some relates to the non-pulsatility of these devices. In those patients for whom the VAD provides a large proportion of their functional cardiac output, their circulatory systems will see only a constant pressure and flow rate, and elastic tissues

[1] ReliantHeart Inc., Houston, Texas, USA
[2] CircuLite Inc., Saddle Brook, New Jersey, USA
[3] Procyrion, Inc., Houston, Texas, USA

within their major vessels will be held at constant tension. Their aortic arch and carotid baroreceptors[102] will become desensitised, as may other as-yet-unknown regulatory systems which are expecting cyclical alterations in blood pressure.

The concern is that long-term patients on constant-flow 'destination therapy' VADs may slowly decline due to failure of homeostatic mechanisms rather than device failures. There is evidence of this in some reports of lower rates of left ventricular recovery with continuous flow versus pulsatile LVADs, but the mechanisms are not yet fully understood[103].

Unlike all the previous pump types discussed in this section – peristaltic or centrifugal pumps for cardiopulmonary bypass, and diaphragm-based positive displacement devices used as first-generation LVADs – the turbine-based axial flow LVADs come the closest to matching the requirements for long-term extracorporeal organ support, yet are also the most technically difficult to design *de novo*.

Advantages & Disadvantages of Axial Flow Pumps

The major advantages of these newer turbine-type flow pumps include:

- The mechanism of action prevents positive and negative overpressures
- Small size, including percutaneous devices under development
- Low energy consumption
- Counterintuitively low rates of hæmolysis and thromboembolisation

All of which result in miniature, portable devices which change the lives of suitable recipients.

The disadvantages however are shared with traditional centrifugal pumps, namely:

- Magnetically-coupled impellers which may decouple from the desired angular velocity if they are required to move faster or slower than their design range
- Flow through the pump is strongly dependent on afterload, such that centrifugal pumps are less able to generate large pressure heads than positive-displacement pumps
- Centrifugal pumps are non-occlusive (non-valved) such that it is possible to get backflow through the device

3.4.3 Summary of Blood Pumping Technologies

In brief, only a small range of pumping mechanisms have translated over from historic industrial designs into clinical blood pumping. These include both positive displacement and dynamic types of pump.

To date the requirements of blood pumps fall into two major categories – cardiac output replacement or supplementation.

Cardiac output replacement occurs in the context of either:

(i) Cardiothoracic procedures not longer than 12 hours, where both peristaltic and centrifugal pump types are able to meet the requirements for cardiopulmonary bypass – flow rates up to 7L/min, with pressures heads of 500mmHg.

(ii) Terminal heart failure where the recipient is placed onto pump support as a temporary (1 – 2 months) 'bridge-to-transplantation' or as a (>6 months) 'destination therapy'. Both pulsatile air-driven diaphragm pumps and now (more commonly) axial flow pumps are able to meet the systemic requirements of 5L/min at 120mmHg. Occasionally centrifugal pumps are placed extracorporeally to act in a 'bridge-to-transplantation' configuration.

Cardiac supplementation is a growing requirement in the developed world in order to treat an ageing population of adults with heart failure (which impacts some 800,000 people in the UK alone[104]).

Most suitable patients now receive either second or third-generation axial flow pumps. These are capable of achieving flow rates up to 2 – 3L/min at 100mmHg pressure – lower pressure and flow requirements than total cardiac replacement, because there is usually some preservation of native ventricular function.

In all these cases, designing pumps to meet the pressure and flow requirements of the human circulatory system is not the major driver of innovation because matching biological pressure/flow requirements are trivial when compared against industrial needs.

Biological pumps must however be:

- Small
- Quiet
- Energy efficient

But most importantly:

- Biocompatible

Therefore, developing a biocompatible blood pump to meet the specific requirements of a single isolated perfused kidney appeared to be a challenge worth addressing.

Chapter 4

Design and Implementation of a Tesla-Type Pump

This Chapter describes the design and implementation of a long-neglected impeller technology which holds theoretical promise for blood pumping with extremely low shear stresses, and thus hæmolysis, owing to its unique mechanism of action – the Tesla pump.

4.1 Invention of the Tesla Turbine

Renowned polymath and pioneer of electrical engineering Nikola Tesla was granted a Patent in 1913 (US1061206A)[105] for his design of "certain new and useful improvements in Rotary Engines and Tubines", which he believed would lead to far simpler and more efficient turbines than those of his day. Figure 4.1 below is taken from this patent.

Tesla recognised that viscous fluids, including air, will adhere to surfaces through boundary effects, creating the property of drag such as that encountered when an object is moved through such a fluid.

Tesla surmised that these boundary effects could be employed to gradually and continuously transfer energy from a fluid to a moving surface, and through the use of an impeller formed from a series of narrowly-spaced stacked parallel discs he sought to create a fully-reversible turbine/pump "free from constraint and disturbance such as occasioned by vanes or kindred devices".

The principles for optimising the disc spacing and angular velocity for given design characteristics were unknown in Tesla's day, and these will be covered later in the Chapter, from Section 4.4 on page 114 onwards.

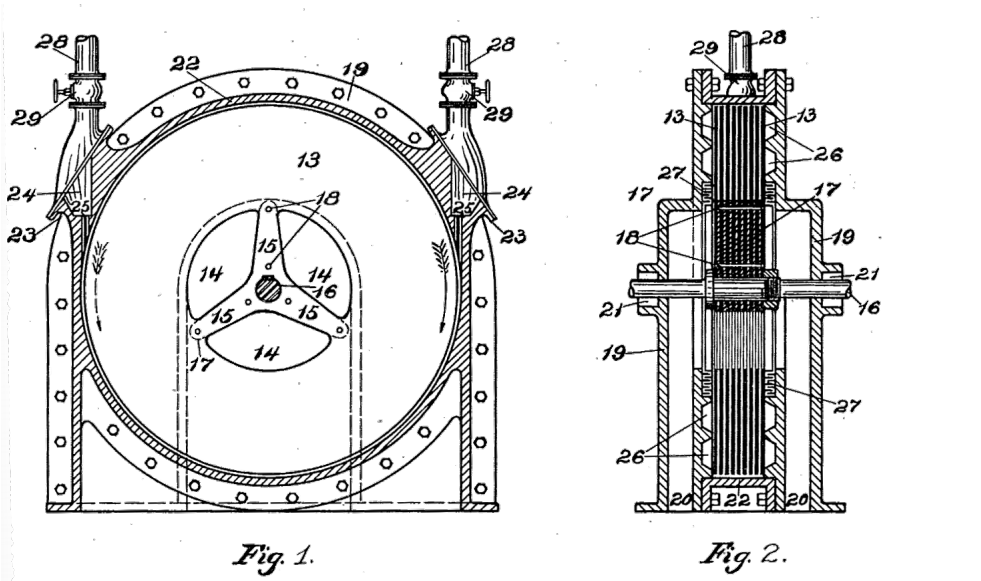


Figure 4.1: Image of Nikola Tesla's Original Design. From Patent US1061206A.

The principle of operation of Tesla-type devices in both pumping and turbine configurations are similar, but for the rest of this Chapter I will only consider the pumping configuration.

4.2 Principle of Operation

A centrifugal pump can be divided into 3 functional parts, as shown in Figure 4.2 below – (i) the inlet, (ii) the impeller, and (iii) the volute leading to the outlet.

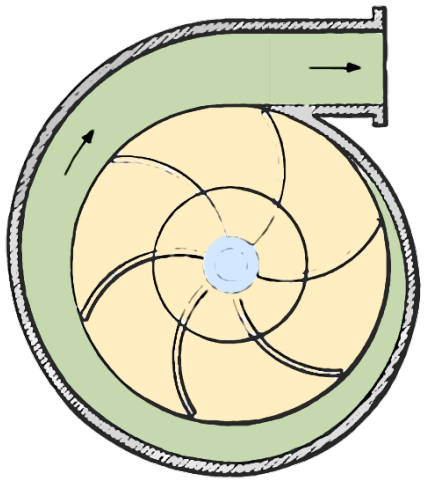


Figure 4.2: Anatomy of a centrifugal pump. **Blue:** The inlet **Orange:** The impeller (a disc stack in a Tesla-type pump) **Green:** The volute

The rest of this Section will consider these three with respect to the unique requirements of the Tesla pump.

The Pump Inlet

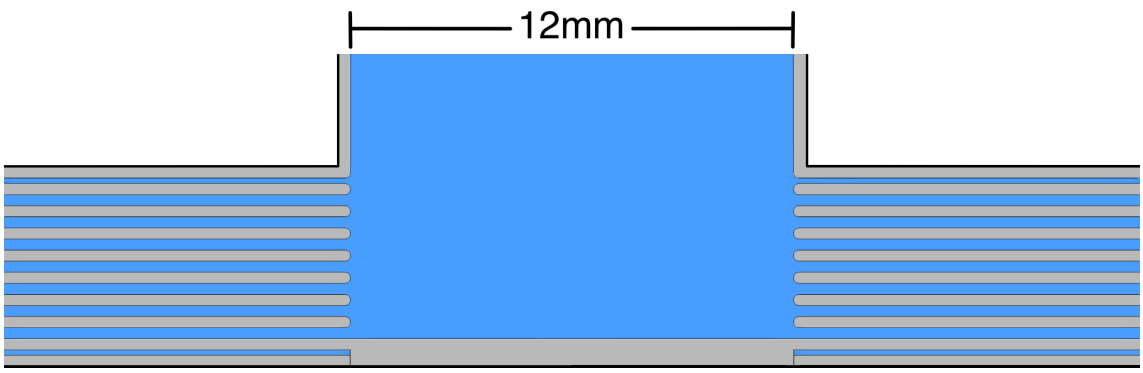


Figure 4.3: Cross-sectional view of the inlet of a Tesla Pump with a 12mm inlet, a solid bottom disc and 300μm inter-disc spacings. Liquid in blue, discs and outer housing in grey. Note the half-height separation between the uppermost and lowermost discs and the housing

Figure 4.3 above shows a cross-section of the inlet portion of a Tesla-type pump. At this stage, the fluid is at a low velocity and low pressure, and the flow will be laminar in nature. The fluid is drawn into the spaces between the parallel plates by the slight pressure drop caused by fluid being propelled away from the inlet boundary through the rotation of the discs. However, this pressure drop is not significant enough to cause cavitation in the absence of pre-load, a problem faced by peristaltic blood pumps.

Early researchers identified this stage of the pump as being a potential source of inefficiency and energy loss[106] as the incoming fluid must be turned through 90° to enter the disc gap. A low inlet flow rate will reduce the magnitude of this energy loss, as may custom stack designs using non-horizontal discs, curved discs or discs with ærofoil-like cross-sections.

Indeed, this entire aspect of the design has remained nascent save for the work of Valentin Izraelev and colleagues based at Pennsylvania State University[107, 108], who examined the effect of altering the horizontal alignment of discs at the inlet stage. Their design is shown in Figure 4.6 on page 114.

The Disc Stack Impeller

Figure 4.4: Rendered image of the packet of stacked discs from my initial Tesla Pump design (page 127). Note the spokes radiating from the central shaft in order to rotate the discs, and the pointed flow divider to redirect liquid into the inlet spaces

In a well-designed disc stack with an appropriate disc-to-disc spacing for the physical properties of the fluid, a pair of boundary layers will develop from each disc and quickly converge to force the fluid into a laminar flow profile. The precise cross-section of the flow profile can be tailored through careful design (see page 115 for the relevant design equations, and my comparison of different algorithms for calculating this optimal space on page 133).

Within this double-thickness boundary layer, the liquid is steadily and constantly

accelerated through a combination of tangential and radial forces. As the tangential force increases with angular velocity and therefore radius, the liquid adopts spiral path lines as it passes outwards through the disc space whilst energy is being imparted.

The greater the angular velocity of the discs, the more radial the flow with accordingly lower rates of energy transfer but greater shear. The lower the angular velocities, the more tangential the flow path with higher rates of energy transfer and lower shear. Therefore the angular velocity of the discs must be tailored to match the physical properties of the liquid against the requirements for low shear stresses whilst producing an adequate pressure head.

Because the flow is forced to remain laminar throughout, no energy is dissipated in turbulence even with a large magnitude change in velocity from the inner radius to the outer radius of the stack.

As discussed on page 100, the BPX-80 Bio-Pump® Plus from Medtronic acts as a boundary-layer pump, whose impeller is comprised of a series of stacked cones. The flow profile that develops between the innermost cones appears to maintain low rates of shear, but the Couette flow between the outermost cone and the inner aspect of the housing imparts large shear stresses[77], resulting in higher rates of hæmolysis than other centrifugal designs[100].

Therefore, limiting the Couette flow component will be an important aspect of a successful design. See page 131 for a discussion of this particular flow profile.

However, the stacked cone design does appear to entrain the blood into the impeller spaces whilst reducing the angle of inlet rotation and energy loss associated with horizontal disc-space inlets. This design would also allow a gravity vector to assist the throughput of the pump, and a cone design may aid in degassing any air emboli from the disc spaces – a phenomenon which leads to severe head loss and flow stalling in the Tesla pump design[106].

The Pump Volute

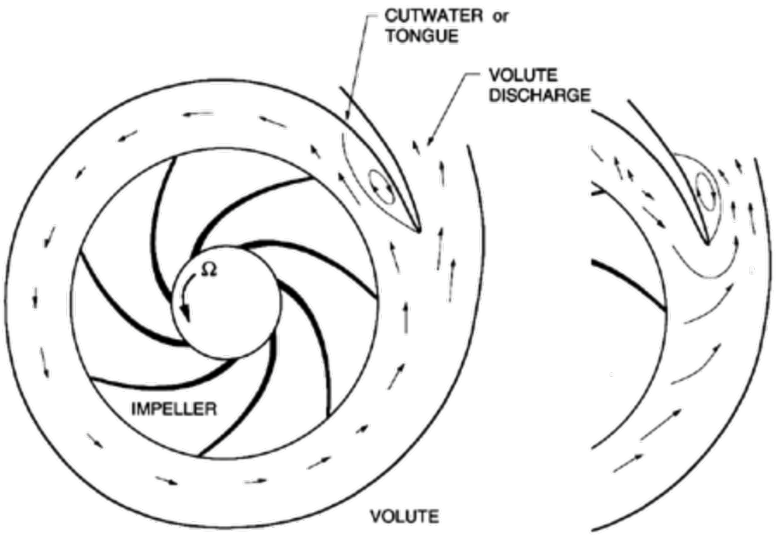


Figure 4.5: Image of a typical volute scroll. Note the appearances of recirculation (left) and cavitation (right) around the 'tongue' with mismatched designs. From [109].

The role of the volute scroll is to convert kinetic energy accumulated in the impeller into a pressure head at the outlet through a steady increase in cross-sectional area. The size and configuration of the scroll, including the location of the 'tongue' can have a large impact on the function of the pump, and is often designed for a narrow working range of flow rates.

The Bernouilli equation describes the total fluid pressure as the sum of the static (the pre-load) and dynamic pressures. The change in kinetic energy of the liquid imparted by the impeller has the following power-law effect on the dynamic pressure:

$$\mathrm{P}_{\mathrm{DYNAMIC}} = \frac{1}{2}\rho \mathrm{V}^2$$

ρ – Fluid density
V – Fluid velocity

At the moment the liquid leaves the outer radius of the disc space, it is acted upon by three forces of unequal magnitude - radial and tangential forces in the horizontal plane, and gravitation in the vertical plane. By tailoring the outer radius and the the angular velocity of the disc stack, the ratio of the radial and tangential forces can be chosen to prefer one over the other. This tailoring may also affect the time spent recirculating in the disc stack, the total path length, and thus the magnitude of energy accumulation within the disc stack.

Flow within the volute should remain laminar so long as the outflow from the disc stack

does not lead to excessive 'spray'-type wall collisions and recirculation, and the placement of the tongue does not lead to cavitation or recirculation.

A volute design suitable for a Tesla-type impeller may indeed be an entire research topic in itself, a factor considered on page 153 at the end of this Chapter.

4.3 Suitability for Blood Pumping

Discussed at length in the previous Chapter, and specifically on page 82, low shear pumps have the potential to resist two of the major negative side effects of blood pumping – hæmolysis and thromboembolisation.

Blood would appear to be particularly well suited to being pumped by a design which relies solely on boundary layer effects for imparting energy, owing to its non-Newtonian characteristics.

4.3.1 Non-Newtonian Characteristics of Blood

Blood is a complex colloid comprising both erythrocytes, with viscoelastic deformable membranes, and charged plasma proteins. This combination causes the blood to behave as a non-Newtonian shear-thinning fluid. [110] provides an excellent review of the state of our understanding of blood rheology. In simplistic terms, at low flow rates, plasma proteins neutralise the negative surface charges on erythrocytes, allowing them to clump and aggregate and increase the dynamic viscosity of the blood. As the flow rate, and therefore the shear rate, increases, the erythrocytes are separated, stretched and flow as individual cells, lowering the blood's dynamic viscosity.

In addition to this shear-thinning effect is the Fåhræus-Lindqvist effect, in which erythrocytes preferentially concentrate at the centre of the channel where the shear rate is lowest, leaving a 'lubricating' low-viscosity plasma layer near the walls. This effect appears in channels less than $\approx 500\mu$m in diameter, and leads to a cell-free layer some 3–4μm in thickness[78, 111].

Therefore, in a well-designed Tesla-type pump, erythrocytes may never be exposed to regions of high shear stress as they are channeled along the centrelines between discs in the stack.

4.3.2 Previous Tesla Blood Pumps

Only a handful of other investigators have explored the possibilities of Tesla-type impellers for blood pumping.

In 1990, a group of researchers at Texas A&M University led by Dr. Gerald E. Miller were the first to publish results demonstrating that this type of pump was capable of matching the performance of a commercial centrifugal blood pump of its day[112]. This initial design comprised 7 polycarbonate discs, each 1.6mm thick and separated by 400μm. The discs each had an inner and outer diameter of 1.5" and 3" respectively. The disc stack rotated at 2250 rpm and generated a theoretical maximum shear stress of 400 dynes/cm ^{2} (40 Pa), lower than that found in normal arteries.

One follow-up report by the same lead investigator in 1999 described a series of experiments performed to reach an optimised design of a 5-disc device with the same disc size but a larger separation of 3.8mm that was able to pump 5L/min at heads of 100mmHg with a rotational speed of just 1,750 rpm[113]. Over a 4-hour period of pumping at this rate, Miller's design resulted in a plasma free hæmoglobin level less than 450mg/L, and below that of the centrifugal pumps against which it was compared[1], and no thrombosis. This value was the equivalent of generating 20mg of free hæmoglobin per 100L pumped.

The work was then seemingly abandoned as no further reports emerged from this group in the literature.

Then, in 2009 and 2011, two papers were published by a group working at the Pennsylvania State University[107, 108] describing a new magnetically suspended Tesla-type LVAD.

CFD modelling identified regions of high shear stress (up to 5000 dynes/cm^2) but isolated to very small regions, which limited the cellular exposure time and thus hæmolysis as per Giersiepen's work[89]. A design iteration led to improved efficiency by revising the disc inlets into a progressively tapered conformation, decreasing the inner disc radius with distance from the inflow, as shown below in Figure 4.6.

[1]Which included the Bio-Medicus™ BP-80 Bio-Pump® – the precursor to the Medtronic BPX-80 Bio-Pump® Plus

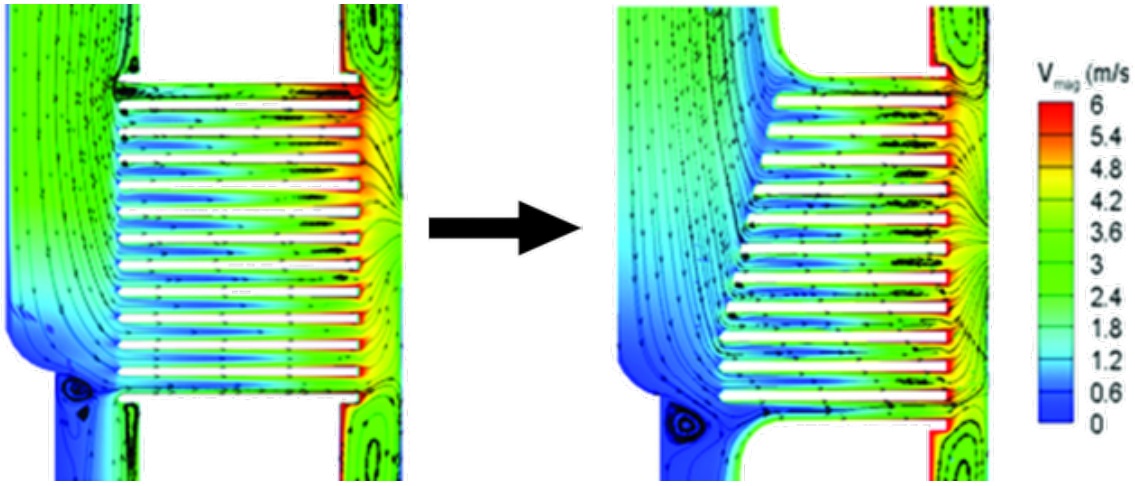


Figure 4.6: Image of Izraelev *et al.*'s tapered inlet design for their Tesla pump, improving throughflow but not preventing regions of flow recirculation. The inlet is to the top left of each image. From [107].

Their final device measured 50mm×75mm, with an outer disc radius of 1cm, a disc thickness of 250μm and a disc separation of 500μm. The device was capable of delivering 6L/min at a head rise of 70mmHg at a rotor speed of 6,750 rpm. Again, the index of hæmolysis was very low at 20mg/100L pumped.

Interestingly, this LVAD used a double-outlet volute design rather than a traditional single-outlet design as is more common with centrifugal pumps. Multi-outlet volutes may in fact prove to be the key to effectively harvesting the output of Tesla-type pumps, an aspect I consider in my conclusions to this work on page 153.

The most recent report by the Pennsylvania State University group dates from 2013 and describes a further iteration of their 2011 design[114]. By reducing the backflow between the lowest disc and the housing, the range of shear stresses is been further reduced from 3,000 – 5,000 dynes/cm^2 to 2,500 – 5,000 dynes/cm^2, with a corresponding drop in hæmolysis rates to 11mg/100L pumped.

These previous investigations reinforced the idea that a Tesla-type pump would be suitable for blood pumping, as it is able to achieve laminar flows at physiological pressures through smooth acceleration of the liquid, resulting in low levels of shear stress and hæmolysis.

4.4 Key Design Equations for Tesla Pumps

The equations governing the design and implementation of Tesla-type pumps have been developed over the past 50 years. Researchers in this field have covered most questions concerning the design and layout of the discs, their spacing and radii, and the angular velocities required to achieve certain output pressures.

Many analyses in the literature, particularly by Rice *et al.*, have focussed on the design in the turbine configuration rather than the pump configuration, and as such the design equations are generalised. This issue is discussed briefly on page 149.

This section gives a brief review of the important historic papers related to Tesla-type pumps, followed by the specifications for a pump to match the biological requirements of a single kidney undergoing normothermic perfusion with an erythrocyte-based solution.

4.4.1 Breiter & Pohlhausen, 1962

Between the publication of Nikola Tesla's original patent in 1913 (US 1061206A)[105] and the work undertaken by Mark C. Breiter and Karl Pohlhausen of the Applied Mathematics Research Branch of the US Aeronautical Research Laboratory in the early 1960s[115], there had been little interest in Tesla's unique design.

Breiter & Pohlhausen were the first to adapt the Navier-Stokes equations to the flow between two parallel rotating discs (unlike Couette's work which formalised the system between one moving plate against a static one). Their solution also introduced one of the key equations governing the design of a Tesla-type pump – the relationship between the disc separation, angular velocity, and fundamental properties of the fluid in order to maintain the boundary layer between the discs.

Beginning with the derivation of the kinematic viscosity of a fluid (ν) from its density (ρ) and dynamic viscosity (μ):

$$\nu = \frac{\mu}{\rho} \tag{4.1}$$

ν – Kinematic viscosity of the fluid between the plates ($m^2 \cdot s^{-1}$)
μ – Dynamic viscosity of the fluid between the plates (*centipoise* (cP), $Pa \cdot s$, or $N \cdot s \cdot m^{-2}$)
ρ – Fluid density (kg/m^3)

Breiter & Pohlhausen showed that the boundary layer thickness between the two discs is related to the square root of the ratio of the angular velocity (ω) and the kinematic viscosity (ν) of the fluid. In order for the disc space (δ) to be occupied be a single fully-developed boundary layer, each disc must contribute one half of the full thickness of the boundary layer:

$$\frac{\delta}{2} \cdot \sqrt{\frac{\omega}{\nu}} \tag{4.2}$$

This non dimensional parameter is also known as the *Ekman number* (E_k)[1], and is crucial

[1]Named after Vagn Walfrid Ekman (1874 – 1954) who studied the movement of icebergs in the Arctic for his

for defining the flow profile between two co-rotating discs.

Simulations in their original paper showed that an E_k of $\frac{\pi}{2}$ gives a fully-laminar plug-like flow profile. This allows us to select an optimal disc separation (δ) for achieving this flow profile based on the kinematic viscosity of the fluid (ν) and the desired angular velocity of the discs (ω).

$$E_k = \frac{\pi}{2}$$
$$\frac{\pi}{2} = \frac{\delta}{2} \cdot \sqrt{\frac{\omega}{\nu}}$$
$$\delta = \pi \cdot \sqrt{\frac{\nu}{\omega}}$$

4.4.2 Hasinger & Kehrt, 1963

The following year, 2 more researchers at the US Aeronautical Research Laboratory published their analysis of the Tesla-type pump[106]. Interestingly, they demonstrated that the flow profile can be manipulated between discs of fixed separation by altering the angular velocity.

For ω values resulting in less than the ideal E_k of $\frac{\pi}{2}$ for a given disc spacing, the flow assumes a parabolic profile as the energy is concentrated in the centre of the flow stream, and for ω values greater than this, the boundary layer is accelerated beyond the centre of the flow stream, leading to a sinusoidal flow profile with a negative central pressure and possible backflow. Thus, tailoring the contour of the radial and tangential flow profiles in this way plays a key part in designing a Tesla-type pump.

Figure 4.7 below reproduces two of the plots from their paper (which they adapted from the work of Breiter & Pohlhausen), demonstrating these flow profiles.

1902 PhD Thesis

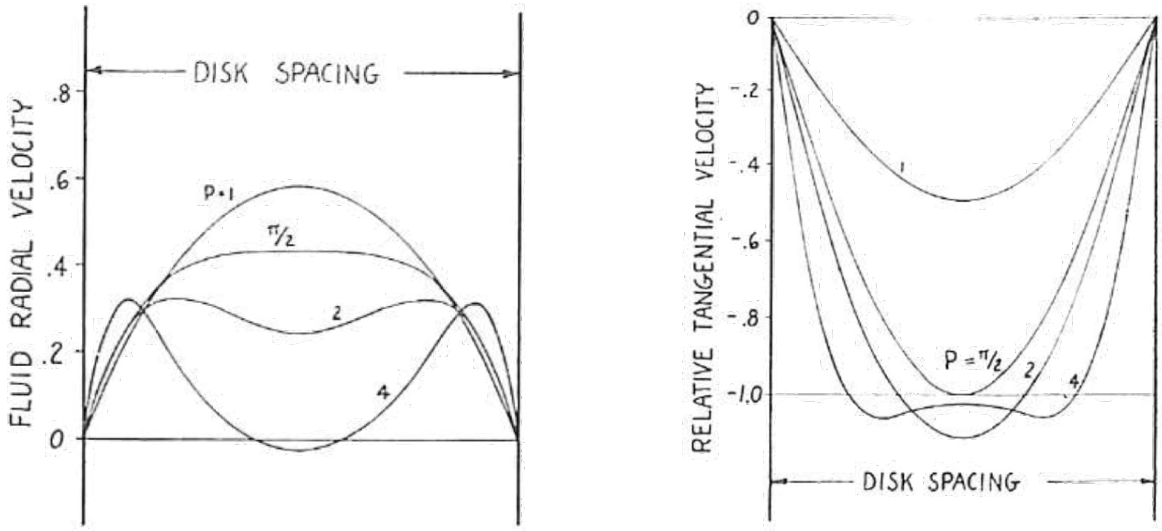


Figure 4.7: Radial and tangential flow profiles between parallel rotating discs vs. the Ekman Number (given here as 'P'). This demonstrates that discs spinning faster than a spacing which allows for adequate capture of the fluid by viscous drag leads to a thin, accelerated boundary layer and midline flow reversal (P = 4). From [106]

They demonstrated that the following two values have large effects on the efficiency of energy transfer by a Tesla pump:

$(\frac{r_i}{r_o})$ – the ratio of the inner and outer disc radii

$(\frac{c_o}{\omega \cdot r_o})$ – The *'flow ratio'* (ϕ) – the radial fluid velocity vs. the outer disc velocity

For good efficiencies and minimal through-flow losses, the authors recommended a $(\frac{r_i}{r_o})$ of less than 0.2, demonstrating that this has a 4th-order effect on energy transfer, and a ϕ value of 0.02, which indicates that the majority of the energy at the outermost radius of the disc is being transferred into tangential movement of the fluid with minimal slippage. They were also the first to consider the need for fluid at the inlet to the disc stack to decelerate through 90° in order to enter the disc spaces and showed that excessive inlet velocities had a drastic effect on head loss.

From the ratio of the disc spacing (δ), the per-space flow rate (q), the kinematic viscosity of the fluid (ν) and the inlet radius (r_i), Hasinger & Kehrt were able to calculate a design chart for ideal Tesla-type pumps giving the dimensionless curve parameter A.

$$A = \frac{\delta q}{\nu {r_i}^2} \tag{4.3}$$

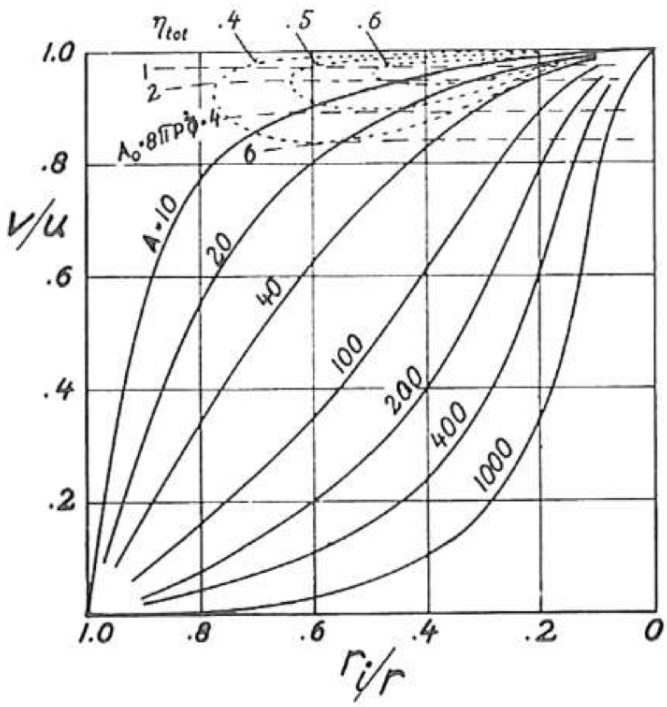


Figure 4.8: Design chart for radial-flow shear-force rotors

The chart in Figure 4.8 above shows the relationship between the ratio of the average fluid tangential velocity (v) to the disc velocity at a given radius (u) as one proceeds from the inlet ($\frac{r_i}{r} = 1$) to the outer disc edge along a given design line (A). These curves describe the rate at which the shear forces transfer kinetic energy from the disc to the fluid at any given radius along the disc.

It is clear that the energy transfer becomes more efficient as the fluid proceeds radially outwards from the inlet, and that one should choose A values below 40 to reduce losses.

One further design chart from the paper demonstrates that to achieve theoretical transfer efficiencies above 60%, one must aim for A values of 5 or less. This chart is reproduced overleaf in Figure 4.9.

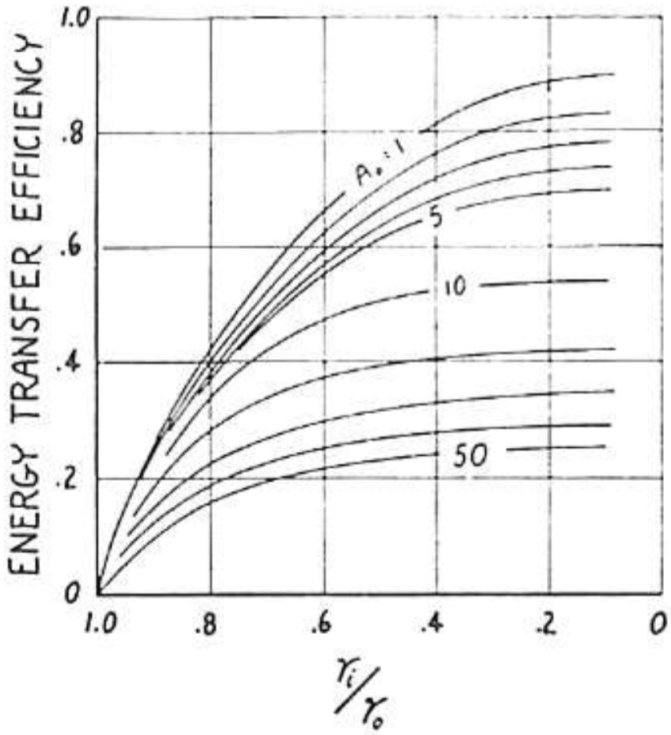


Figure 4.9: Theoretical transfer efficiencies of radial-flow shear-force rotors

Hasinger & Kehrt also simplified Breiter & Pohlhausen's equation for the total flow through the pump to:

$$Q = 2\pi r_o \cdot c_o \cdot \delta \cdot z \tag{4.4}$$

Q – Total pump flow
c_o – Radial flow at the outer radius (r_o) of the pump
δ – Disc spacing
z – Total number of disc-spaces (where the total number of discs = $z + 1$)

However, their analysis did not formally explore the role of the Couette flows which develop between the uppermost and the lowermost discs and the stationary housing, and instead put these losses down to general 'friction losses' in their experiments.

4.4.3 Matsch & Rice, 1968

Professor Warren Rice of Arizona State University pursued research into Tesla-type turbomachinery from the late 1960s onwards. However, his work focussed mainly on turbine configurations rather than pumps, a difference considered later on page 149. One important factor his work has defined and investigated is an *'operating Reynolds number'*,

N_{Re}[116], which is comprised of the fundamental properties of the fluid and the physical design of the turbine.

$$N_{Re} = \frac{\omega\delta^2}{\nu} \tag{4.5}$$

Rice's 1974 paper[117] contains calculated design data for Tesla-type turbines using incompressible fluids. Note that because of the dimensionless nature ofN_{Re}, these results should still be applicable to Tesla-type pumps as well as Tesla-type turbines.

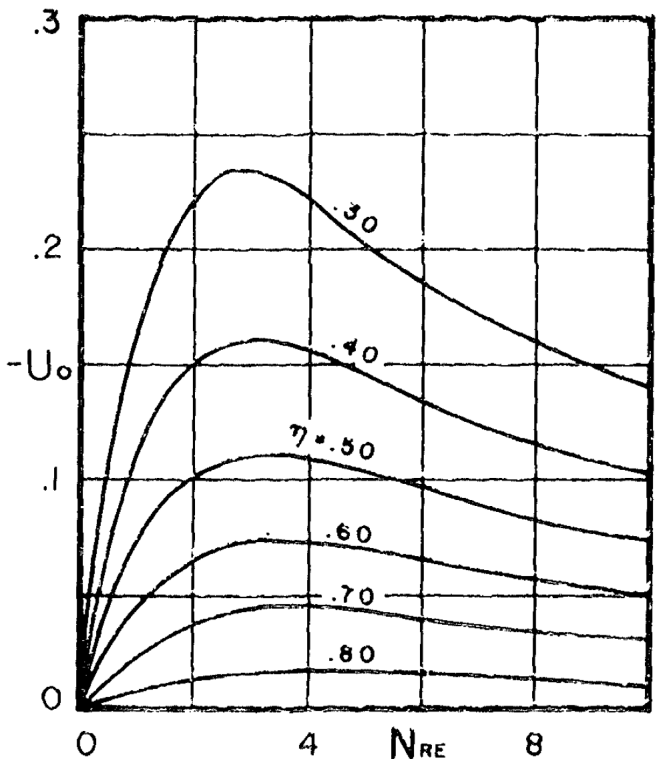


Figure 4.10: Constant efficiency lines for N_{Re} vs. the average radial velocity at entrance (U_o)

Figure 4.10 is just one of many from the paper, each calculated for different values of radial inlet velocity (U_o) and an ($\frac{r_i}{r_o}$) ratio of 0.3 – below the optimal ratio determined by Hasinger & Kehrt.

Interestingly, each of the curves demonstrates peak efficiency when N_{Re} lies between 2 – 4, and that lowering the inlet radial velocity towards stasis improves the overall pump efficiency, confirming the earlier findings of Hasinger & Kehrt regarding inlet head loss due to the angle through which the liquid must decelerate in order to enter the disc gap.

4.4.4 Oliveira and Páscoa, 2009

Publications about Tesla-type pumps seem to have slumped during the 1980s and 1990s. However, the 2000s saw a new interest in this low component count, low flow rate, low Reynolds number pumping method for potential *MEMS* applications.

A presentation at the 3rd national conference on fluid mechanics in Bragança, Portugal in 2009 by Oliveira & Pascoa provides an insight into the design methodology for a customised Tesla-type pump[118] using magnets embedded in the discs for momentum transfer.

They also derive the angular velocity required to achieve a desired total pressure rise from the pump from a simplified equation for the stagnation pressure, as well as making the observation that efficiency increases up to a maximum $(\frac{r_i}{r_o})$ ratio of 0.1.

$$\omega = \sqrt{\frac{2P_t}{\rho(2r_o^2 - r_i^2)}} \tag{4.6}$$

ω – Angular velocity
P_t – Total pressure rise
ρ – Fluid density
r_o – Outer disc radius
r_o – Inner disc radius

4.5 Summary of Useful Design Equations

The important equations for designing a Tesla-type pump with a plug-like radial flow profile are listed in the Table 4.1 below. These are used later in this Chapter to produce my Tesla pump designs as demonstrated in Section 4.8 (page 126) onwards.

Parameter	Equation
Boundary layer thickness (δ) for an Ekman number (E_k) of $\frac{\pi}{2}$	$\delta = \pi \cdot \sqrt{\frac{\nu}{\omega}}$
Optimum Operating Reynolds number (N_{Re})	$2 \leq \frac{\omega\delta^2}{\nu} \leq 4$
Angular Velocity (ω) for a given target Pressure (P_t)	$\omega = \sqrt{\frac{2P_t}{\rho(2{r_o}^2 - {r_i}^2)}}$
Optimal Curve Parameter (A) from Hasinger & Kehrt	$0 \geq \frac{\delta q}{\nu {r_i}^2} \leq 10$
Optimal 'flow ratio' (ϕ)	$\frac{c_o}{\omega r_o} \leq 0.02$
Total stack flow rate (Q)	$Q = 2\pi r_o \cdot c_o \cdot \delta \cdot z$

Table 4.1: Key Equations for designing a Tesla pump

The other parameters required for designing a pump to meet the specific physiological pressure, flow and oxygen requirements of a single isolated perfused kidney are derived from the biology of the organ and blood rheology.

4.6 Relevant Biological Data

Meeting Renal Oxygen Requirements

This topic was first introduced on page 61 as part of a larger discussion on renal metabolism (Section 2.5). This Section builds our understanding of the biology of renal oxygen consumption into design parameters for a blood pump.

4.6.1 Erythrocyte-Based Perfusate Calculations

In the case of an isolated perfused system, the designer has far more control over the components of the perfusate than for pumps required for whole blood. It is therefore possible to resist thromboembolisation by removing all components of the clotting cascade from the perfusate and use only erythrocytes in an isotonic electrolyte suspension with

some plasma proteins added for stability. Lowering the erythrocyte content also results in a lower likelihood of erythrocyte damage and hæmolysis.

Such plasma-free erythrocyte-based perfusates have already been proven in an isolated warm machine perfusion paradigm[34]. There is also evidence from studies undertaken through the 1980s up until 2003 that the addition of small quantities of macromolecular polyethylene glycols with molecular weights from 1kDa – 20kDa to erythrocyte suspensions are protective against mechanical hæmolysis, possibly though membrane-stabilising effects[119]. However, no other groups seem to have continued this work with circulating PEGs since, and they are not used in clinical practice, so they will not be considered further here.

Here we will derive the physical characteristics of an erythrocyte suspension specifically designed for perfusing a single kidney.

Calculations

As originally discussed in Section 2.5 on page 61, a single kidney receives approximately 500ml of blood per minute from the heart, and from this it extracts at most 500μmol/minute of O_2[19, 56], equal to a maximum O_2 consumption rate of 11.2ml/min at STP.

Because hæmoglobin has an oxygen-binding capacity of 1.34ml of O_2 per gram (Hüffner's Constant), the kidney must receive 8.5g of fully-saturated hæmoglobin per minute to satisfy its oxygen requirements. At a flow rate of 500ml/minute, this requires a minimum hæmoglobin concentration of 17g/L in the recirculating liquid.

However, because we cannot expect to totally extract all of this oxygen from the hæmoglobin flowing through the kidney, we can use a large safety factor of 200% to arrive at final hæmoglobin concentration of $\sim$70g/L, still under half the normal adult male hæmoglobin concentration of 150g/L, and equivalent to a hæmatocrit[1] of 20%.

4.6.2 Blood Viscosity (μ)

The equation relating dynamic viscosity (μ) to hæmatocrit gives the following result for a hæmatocrit of 20% (Zingg *et al.* [120]):

[1]Percentage of erythrocytes comprising the total blood volume

$$\mu = 1.24e^{0.02471 \cdot \text{Hct } \%}$$
$$\mu = 1.24e^{0.02471 \cdot 20}$$
$$\mu = 2.033 \text{ centiPoise (mPa} \cdot \text{sec)}$$

Note that normal blood has a dynamic viscosity on the order of 3.5 centiPoise.

4.6.3 Blood Density (ρ)

Because we will be using a solution of washed erythrocytes suspended in isotonic solution with only a low protein content, with the hæmatocrit known, the density (ρ) can be calculated using the known density of water at 37°C and the density of red cells as measured by Grover *et al.*[121].

$$\rho_{H_2O} = 993.3 \text{ kg/m}^3$$
$$\rho_{RBC} = 1110 \text{ kg/m}^3$$
$$\rho_{20\% \text{ Hct}} = (80\% \times \rho_{H_2O}) + (20\% \times \rho_{RBC})$$
$$\rho_{20\% \text{ Hct}} = 1015 \text{ kg/m}^3$$

Note that normal blood has a density on the order of 1060 kg/m^3.

4.6.4 Inlet Diameter

A cadaveric study by Satyapal *et al.* in 1995[122] determined that the average human renal vein measures 1.2cm $\pm$0.2cm in diameter. The renal arteries are on the order of 6mm $\pm$1mm in size.

This is the final piece of information required to design a Tesla-type pump to match the flow, oxygen delivery and pressure requirements of an isolated perfused human kidney.

4.7 Final Design Considerations

The final stage before commencing the design stage is to enumerate the key design constraints.

(i) The pump must not cause hæmolysis through shear stress
(ii) The pump should be able to generate either a constant or pulsatile flow
(iii) The pump must be small enough to fit in a portable device, and be ideally no larger than the kidney itself

The first of these should be 'guaranteed' by the use of a Tesla pump, the second by calculation of the boundary layer flow profile for a range of outlet pressures.

The third of these constraints is less well defined.

There are a great number of degrees of freedom when customising the design of a Tesla pump, but it given the evidence it would seem prudent to constrain the $(\frac{r_i}{r_o})$ ratio to $\simeq$0.1 as per the findings of Hasinger & Kehrt. With an inlet radius equal to that of the renal vein, this results in a disc stack between 10 – 12cm in diameter. Note that the total diameter of the pump will be on the order of a few centimetres greater than this when the volute is taken into account.

Much of the research from the 1960s to the present has focussed on optimising the performance of Tesla turbines in terms of efficiency of energy transfer. A few more recent papers have begun to address the design constraints in terms of miniaturisation for blood pumping as LVAD devices[107, 108], but there are additional real-world constraints to be considered at this stage of the project.

4.7.1 Manufacturability

The most important real-world constraint is the manufacturability of a practical design, and balancing the degree of precision desired from the design against an appreciation of the likely inaccuracies when translating the equations into a physical device.

While 3D printing allows a wide degree of customisation for the disc housing, it is not yet sufficiently advanced to allow the printing of pre-assembled large aspect-ratio disc stacks to micron tolerance, and so we must turn to traditional prefab sheet materials instead - steel and plastics.

Steel sheets less than 1mm in thickness are usually available in step sizes of 100μm with tolerances of $\pm$5μm (data from various manufacturer and supplier websites).

Polycarbonate sheets should also be sufficiently rigid to produce a useable disc, and there are commercially-available polycarbonate sheets in step sizes of 100μm from 200μm to 1mm[1]

We can then assume the same thickness would be required for any washers or spacers made the same materials. Any disc-to-disc spacing mechanism produced from 3D printed material should be accurate to 50μm.

This is therefore the greatest design constraint, limiting our ability to customise the disc-to-disc spacing (δ) beyond 50μm intervals.

Using a laser (or other) cutter to produce disc shapes from a sheet of steel or acrylic should allow us to select an outer radius accurate to 100μm[2]

Finally, the number of discs to be used in the disc stack must be an integer.

4.8 Design Constraints

The following list summarises the fixed constraints of the design.

- Dynamic Viscosity (μ) = 2.033cP (0.02033 g/cm·s)
- Fluid Density (ρ) = 1015kg/m^3
- Inner Radius (r_i) = 0.6cm
- Radius Ratio ($\frac{r_i}{r_o}$) = 0.1 – 0.2
- Disc Spacing (δ) = 200μm - 500μm in 50μm steps
- Pressure Head (P_t) = 100mmHg + 20% *'safety factor'* to account for pump losses
- Volume Flow (Q) = 500ml/min

Calculated Parameters

And the parameters we wish to optimise in the final design:

- Number of Discs
- Disc Spacing (δ)
- Angular Velocity (ω)
- Shear Stresses to limit hæmolysis

Taking these considerations in mind, it is now possible to explore possible designs for the disc stack.

[1]Hesa®-Glas Plastic Sheets, Peerless Plastics and Coatings Ltd, UK.
[2]Laser cutters are said to be accurate to $\pm 10\mu$m

4.9 Initial Design and Testing

Calculations based on the design equations previously covered in this Chapter and summarised on Table 4.1 (page 122) were performed using custom software written in Matlab R2012b on a 2013 Apple MacBook Pro 2.7 GHz Core i7.

This design began with the thinnest steel sheeting available and a desire to create a disc stack no larger than 5cm in diameter (25mm radius). This 254μm sheet was used to create washers to separate 500μm-thick steel discs with an outer radius of 25mm and an inner radius of 6mm. An image of this first disc stack was shown in Figure 4.4 on page 109.

Three spokes were used attach each disc to the central spindle. Accommodating the decrease in inlet surface area as a result of these spokes led to a compensatory increase in the inner radius to 8mm.

4.9.1 Calculations

We begin by calculating the angular velocity for a target pressure head of 120mmHg, given an r_o value of 6cm, and an r_i of 0.6cm, using the equation of Oliveira and Páscoa (equation 4.6), given the values for the inner and outer radii derived above:

$$\begin{aligned}
\omega &= \sqrt{\frac{2P_t}{\rho(2{r_o}^2 - {r_i}^2)}} \\
\omega &= \sqrt{\frac{2 \cdot 16,000\text{Pa}}{1015 \cdot (2 \cdot 0.025^2 - 0.008^2)}} \\
\omega &= 163.03 \text{ rad/sec} \\
&\simeq 1540 \text{ rpm}
\end{aligned}$$

Followed by a calculation of the number of discs required according to Hasinger & Kehrt (equation 4.4) for a 'flow ratio' (ϕ) of 0.02 with a safety margin of 30% extra flow (650ml/min):

$$Q = 2\pi r_o \cdot c_o \cdot \delta \cdot z$$

Where:

$$\phi = \frac{c_o}{\omega r_o}$$

Therefore:

$$
\begin{aligned}
Q &= 2\pi r_o^2 \cdot \phi \cdot \omega \cdot \delta \cdot z \\
650\text{ml/min} &= 0.04\pi \cdot 2.5\text{cm}^2 \cdot 163.03\text{ rad/s} \cdot 254\mu\text{m} \cdot z \\
z &= 10.4
\end{aligned}
$$

Rounding to the nearest integer leads to 11 disc spaces, for a total of 12 discs.

4.9.2 Final Design Data

μ = 2.033cP
ρ = 1015kg/m^3
r_i = 0.8cm
r_o = 2.5cm
P_t = 120mmHg (including the 20% *'safety factor'* for pump losses)
ω = 1540 rpm (163.03 rad/s)
δ = 254μm
Disc Thickness = 500μm
Disc Number = 12

4.9.3 Pump Manufacture

The pump was designed in SolidWorks® Education Edition 2013–2014[1] on a 2013 Apple MacBook Pro 2.7 GHz Core i7 running Windows 7 in Boot Camp.

The disc stack was designed according to the parameters listed above, with a conical cap sitting atop the central shaft to act as a flow director. It was driven by a small brushless 11.4V DC motor[2] designed for remote-controlled aircraft, with its speed under manual control.

The inlet and outlets were designed with simple friction fittings to interface with external piping, and the volute was designed as a simple spiral beginning at a radius 0.5mm greater than r_o and expanding outward to a radius equal to 1cm greater than r_o.

The SolidWorks® design was then built using the department's Objet30™ [3] 3D printer, and assembled with the disc stack and the DC motor for testing. This was printed in two parts: (i) the lid with the inlet, and (ii) the body with standoffs and the outlet. The image

[1] Dassault Systèmes, Vélizy-Villacoublay, France
[2] Grand Turbo GF2210/20, EMax US Inc., Brea, California, USA
[3] Stratasys Ltd., Eden Prairie, Minnesota, USA

of the fully-assembled device formed the image introducing this Section of the Thesis, and derived from the 3D design shown below in Figure 4.11.

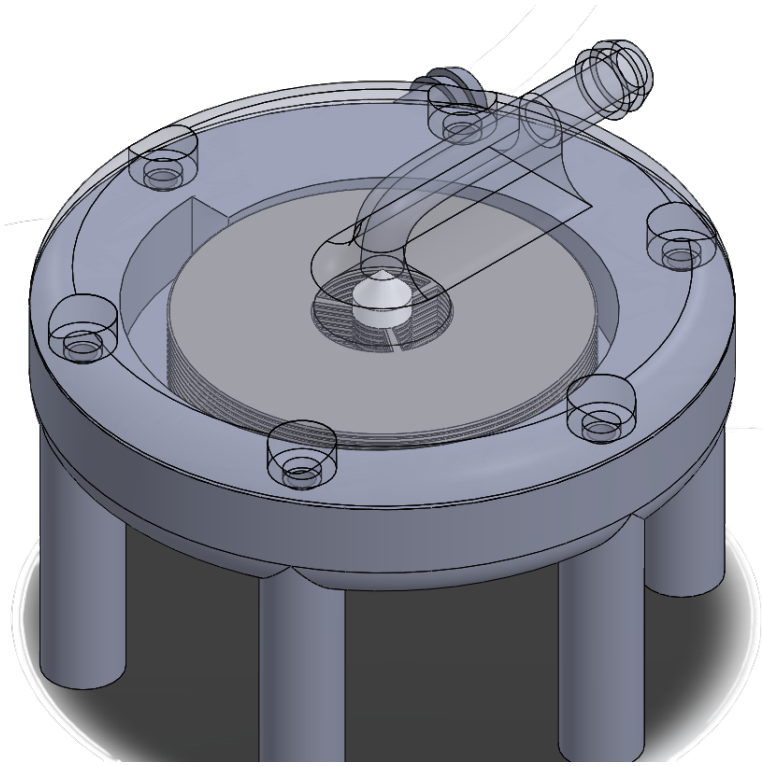


Figure 4.11: Image of the final pump rendered in SolidWorks® demonstrating the lid with the integrated inlet, the white conical cap, the central disc stack with spokes, and the body with the stand-offs to allow the DC motor to fit underneath

4.9.4 Findings and Explanation

The pump performed very poorly in testing. It was unable to self-prime, and following priming it was regularly choked by a large air pocket which collected at the inlet cap. As a result, it was not able to achieve any significant pressure or through-flow at the outlet.

One major problem was the visible recirculation of flow within the pump, made more obvious when food colouring was added to the inlet stream. Whether this was as a result of the macroscopic entrapment of air at the inlet, or part of the same problem, was unknown.

These problems are demonstrated more clearly in the video on the disc accompanying this Thesis, from which the sequence of still images in Figure 4.12 were taken.

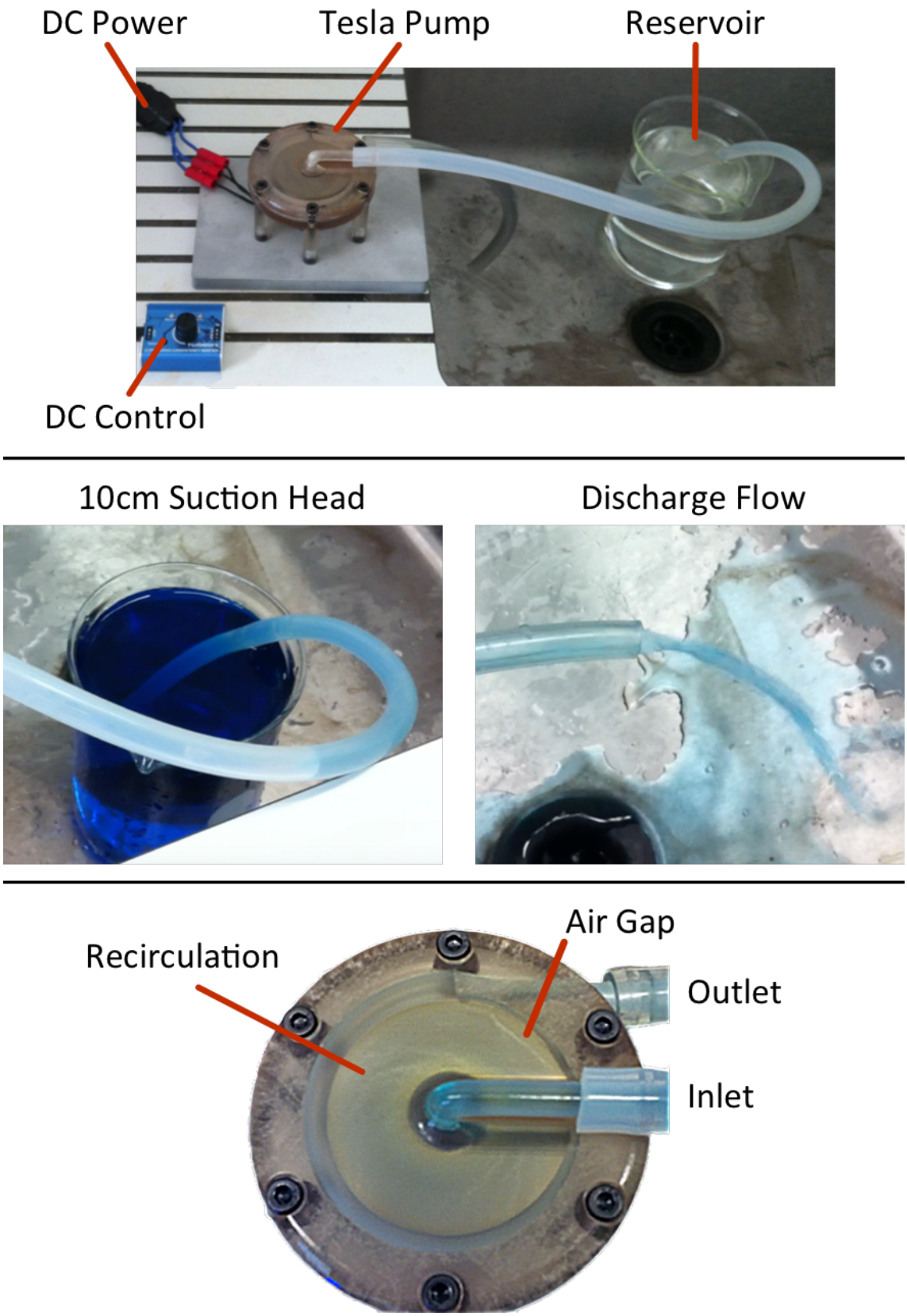


Figure 4.12: Still images extracted from the video of the first prototype pump undergoing testing **Top:** Experimental setup **Middle Left:** The pump is unable to self-prime **Middle Right:** Low pressure outflow **Bottom:** Recirculation and development of an air gap beyond the tongue of the outlet

It was believed that problems were likely a combination of manufacturing error and/or inadequate volute design to match the intended performance of the disc stack, and so a basic fluid dynamics simulation was performed in SolidWorks® to try to elucidate the problem, with the inlet and outlet flow rates fixed at 500ml/min.

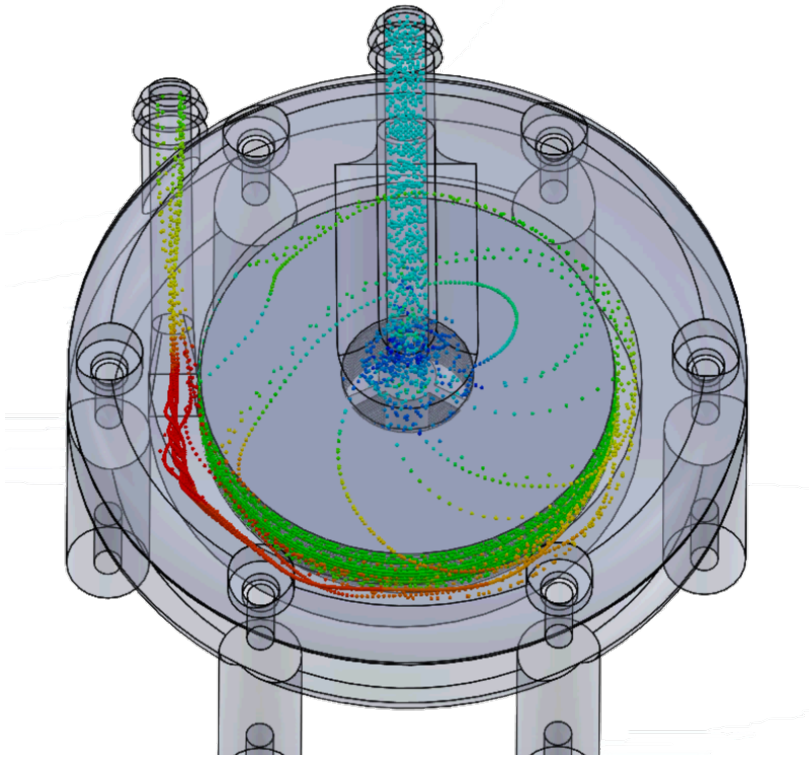


Figure 4.13: Particle-based fluid dynamic simulation in SolidWorks® demonstrating the problem of recirculation from multiple points around the periphery

This simulation demonstrated the recirculation problem quite clearly. Fluid is initially entrained between the disc spaces, but as it is ejected peripherally it becomes constricted near the outlet (red particles indicating higher pressures) and forced to recirculate. There are also multiple peripheral sources of recirculation back into the region of the inlet from the periphery.

Couette Backflow

The most likely explanation for these results is that the pressure differential from the high velocity fluid exiting the discs and the near zero velocity at the inlet favours a reverse Couette flow profile between the uppermost rotating disc and the stationary housing, as shown in Figure 4.14 below. In this initial pump design, the clearance between the upper disc and the housing was 1.2mm, some 5× the disc separation, strongly favouring this flow profile and leading to a 'leaky' design. This was an important finding avoided in later designs.

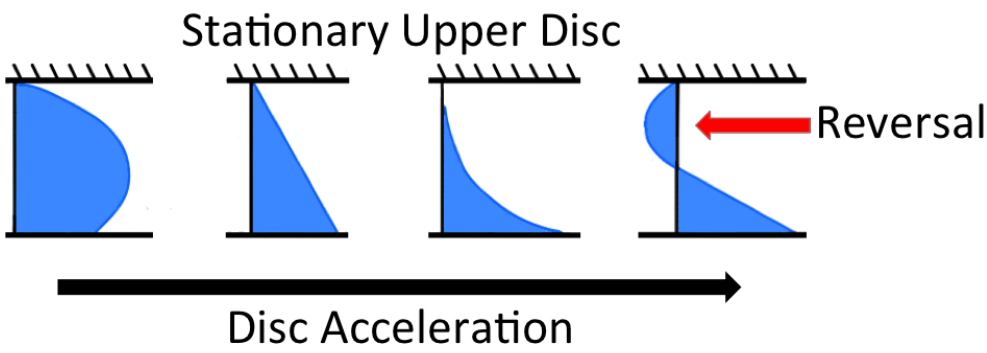


Figure 4.14: Evolution of the flow profile as the disc accelerates, until flow reverses near the stationary upper surface. The classical Couette profile is the second from the left

4.10 Second Pump Design

It now appears that the correct starting point for any Tesla pump design is to minimise the $(\frac{r_i}{r_o})$ ratio, followed by optimising for the flow profile between the discs. For this reason I began by increasing the outer diameter from 5cm to 6cm.

An attempt was also made to minimise the Couette backflows by reducing the spacing between the disc stack and the housing to $(\frac{\delta}{2})$ and by making the bottom disc solid, imagining that it may instead be driven by a magnetic spindle lying just beneath the housing.

4.10.1 Calculations

We begin again by calculating the angular velocity for a target pressure head of 120mmHg from Oliveira and Páscoa[118]:

$$\omega = \sqrt{\frac{2P_t}{\rho(2{r_o}^2 - {r_i}^2)}}$$

$$\omega = \sqrt{\frac{2 \cdot 16,000\text{Pa}}{1015 \cdot (2 \cdot 0.06^2 - 0.006^2)}}$$

$$\omega = 66.3356 \text{ rad/sec}$$

$$\simeq 630 \text{ rpm}$$

From this point is is possible to derive the optimal disc spacing from one of two previously-established design equations:

- Breiter & Pohlhausen's equation, and an E_K of $\frac{\pi}{2}$ from Hasinger & Kehrt's flow profile charts
- Matsch & Rice's requirement for the optimal *operating Reynolds number* (N_{Re}) to lie between 2 – 4

It was therefore decided to compare the results of each disc spacing in simulation to try to determine the optimal result.

4.10.2 Comparing Optimal Disc Spacings

Breiter & Pohlhausen's equation gives the following disc spacing (δ) for an E_K of $\frac{\pi}{2}$:

$$\delta = \pi \cdot \sqrt{\frac{\nu}{\omega}}$$
$$\delta = \pi \cdot \sqrt{\frac{\mu}{\omega \cdot \rho}}$$
$$\delta = \pi \cdot \sqrt{\frac{0.002033 \text{ kg/m} \cdot \text{s}}{66.3356 \text{ rad/s} \cdot 1015 \text{ kg/m}^3}}$$
$$\delta = 546\mu\text{m}$$

Matsch & Rice's equation gives the following disc spacing (δ) for an N_{Re} of π:

$$\pi = \frac{\omega\delta^2}{\nu}$$
$$\delta = \sqrt{\frac{\pi \cdot \nu}{\omega}}$$
$$\delta = \sqrt{\frac{\pi \cdot \mu}{\omega \cdot \rho}}$$
$$\delta = \sqrt{\frac{\pi \cdot 0.002033 \text{ kg/m}}{66.3356 \text{ rad/s} \cdot 1015 \text{ kg/m}^3}}$$
$$\delta = 308\mu\text{m}$$

Back-calculating the Ekman number from this value of δ gives a value of 0.8862, indicating that the radial flow profile should be highly parabolic.

4.10.3 Simulation in COMSOL

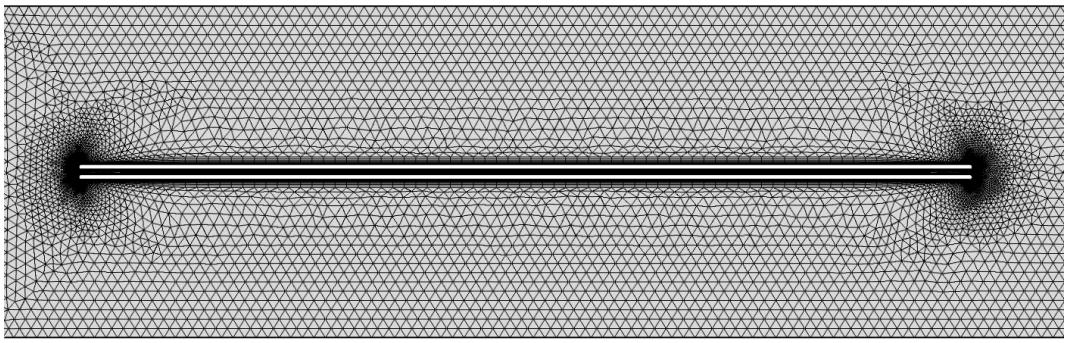


Figure 4.15: 2D axisymmetric simulation of the two discs spinning freely in the simulated erythrocyte-based perfusate. Note that the mesh density increases around the inlet and outlet as well as in the thin channel to optimise the simulation

A series of 2D simulations were undertaken using COMSOL Multiphysics[1], a finite-element software package for engineering applications, using the values calculated by the Matlab program.

2D Axisymmetric Simulation

Rather than constructing the entire model in 3D, a simplified 2D axisymmetric model was created using COMSOL's internal editor.

This modelled the disc pair spinning freely within a simulated volume of the erythrocyte-based perfusate 3× larger than the disc outer radius and under the effect of gravity. The 2D geometry did not include the central rotor shaft or any struts or disc-to-disc support structures that would maintain the calculated disc spacing. Note that the disc spacing and disc thickness were equal in these models, unlike in the first prototype where the disc thickness was double that of the disc spacing.

A custom incompressible Newtonian fluid was created using the viscosity and density values previously calculated for blood at a 20% hæmatocrit. Unfortunately, real blood is a non-Newtonian fluid but simulating this was beyond the capabilities of the COMSOL software.

The appropriate boundary conditions were set to meet the pre-determined target angular velocity of the discs, with the gauge pressure fixed at 0 Pa in the region of the inlet.

[1]COMSOL Multiphysics Version 5.0, COMSOL AB, Stockholm, Sweden

Meshing the Model

The geometry was meshed using COMSOL's internal meshing capabilities, based on settings for fluid dynamics with additional boundary layer refinements to solve the flow profile between the narrowly-spaced discs. Discussions with other researchers indicated a general 'rule of thumb' that simulation of boundary layer effects requires a depth of at least 5 cells growing from the moving surface.

For this reason I chose to use a free triangular mesh with an element size of 26μm – 938μm, and 8 boundary cells from each disc surface with a growth factor of 1.2. This gave a final mesh with 17604 elements, including 958 boundary elements. The result was shown previously in Figure 4.15 on page 134.

The minimum element quality after meshing was deemed adequate by COMSOL's internal standards, with a 'quality factor' of 0.6921, greater than the minimum recommended value of 0.1.

Calculating the Reynolds Number

A calculation of the Reynolds number defined by the disc separation as the reference length and the angular velocity of the tip of the disc gave the following result:

$$\begin{aligned} Re &= \left(\frac{\rho \omega r_o \cdot \delta}{\mu} \right) \\ &\simeq 306 \end{aligned}$$

Indicating that the flow between the discs would be laminar in nature and fall between that of a vein (≈140) and an artery (≈500). The same value without the adjustment for the scale of the disc separation, thereby looking at the value at the rotating disc edges, was on the order of 30,000 which indicated low turbulence. The turbulence intensity (I) in this region was only:

$$\begin{aligned} I &= 0.16 \times Re^{-\frac{1}{8}} \\ &\approx 4\% \end{aligned}$$

Indicating a medium turbulence intensity at the disc tips.

Solving the Model

Given the majority of the flow in the system would be laminar in nature and that previous reports described only small changes with RANS-based simulations for Tesla-type pumps[108], I chose to use a laminar flow model for the simulation for a combination of speed and convenience.

The model was solved as a time-dependent flow with a smooth ramp function applied to the angular velocity of the discs to accelerate them from 0 rad/s to full speed over 1 second. The solution was then continued for 0.1 second after full flow conditions were achieved in order to model the steady state behaviour of the disc packet.

Models were remotely submitted to Imperial College's High Performance Computing Cluster (HPC) for processing on a 20-CPU node with 80GB RAM per CPU. The CPUs were 2.20GHz Intel(R) Xeon(R) E5-2660 v2 (Ivybridge) with DDR3 1866Mhz memory. Solving the models took ≈12 hours of CPU time on these nodes.

Simulation Results

These results explore various aspects of these two different disc spacings in an attempt to minimise shear stress and therefore potential hæmolysis.

Velocity Profiles

The following graphs in Figure 4.16 illustrate the midline velocity profiles for both disc separations at 66.3356 rad/s:

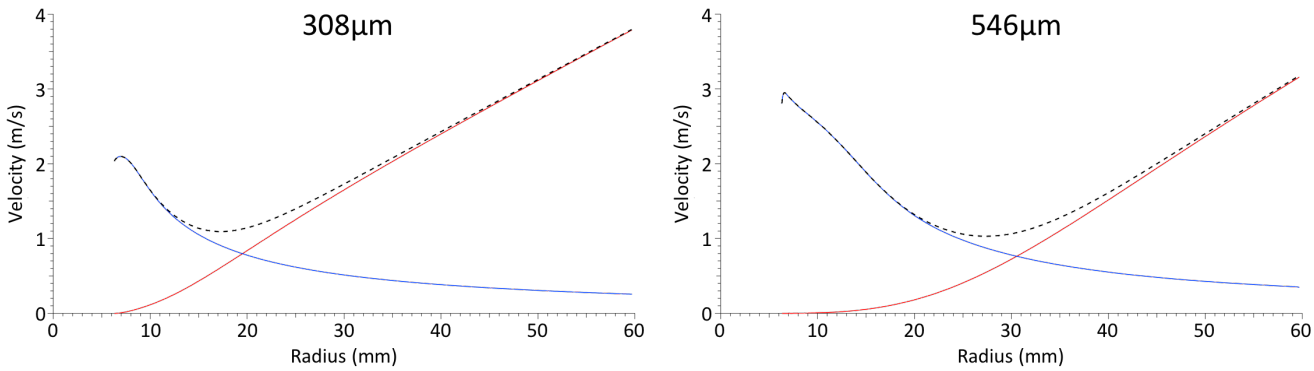


Figure 4.16: Midline flow velocities for 308μm and 546μm disc separations. Red line = tangential velocity, blue line = radial velocity, dotted black line = total velocity

It is very interesting to see that both profiles describe an initial rapid peak in velocity as liquid is entrained at the inlet, followed by a rapid decrease in the radial velocity as the

boundary layer develops along the spinning disc and the liquid begins to move tangentially along with the disc.

The transition point between the majority radial and majority tangential flow occurs at different points for each disc separation - approximately 13mm from the inlet (19mm - 6mm) for the 308μm separation, and 24mm for the 546μm separation.

It appears that the 308μm separation results in a greater outlet velocity than 546μm because of less total energy loss through the transition point, measured as the difference in velocity between the inlet and the nadir of the transition point.

Flow Profiles and Shear Stresses

Figure 4.17 illustrates the development of the radial flow profiles from the inlet, through the transition zone and to the outlet for both disc separations. Note that the 546μm spacing does achieve a plug-flow profile at the outlet as predicted by the design equation. Values are in metres/sec:

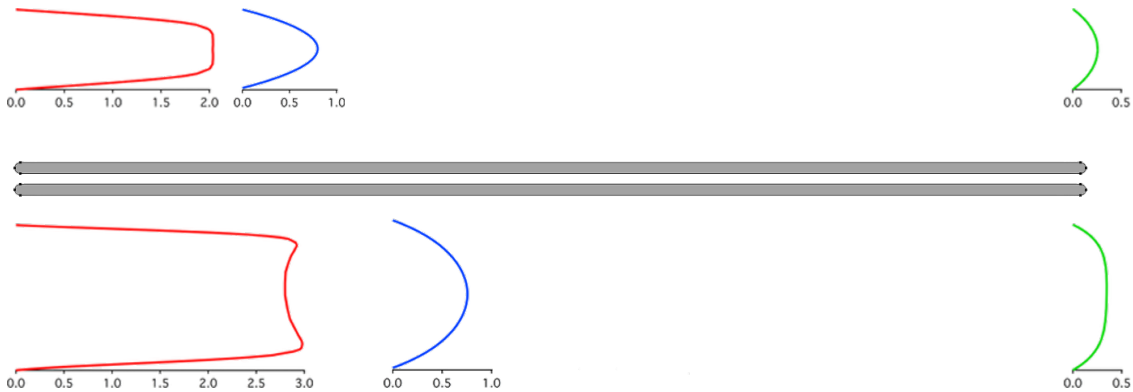


Figure 4.17: Development of the radial flow profiles for δ = 308μm (top) and δ = 546μm (bottom) disc separations

COMSOL is able to calculate the per-cell shear stress (in Pa) from the product of the shear rate (s^{-1}) and the dynamic viscosity (μ) of the sample liquid[1]. Unfortunately it was not able to collect the stresses at the wall due to the way in which the region of interest is selected for analysis within COMSOL, hence the plots in Figure 4.18 illustrate the midline shear stresses experienced radially along the disc.

A future study could therefore aim to extract additional data by analysing the maximal shear stresses along each vertical step progressing from the inlet to the outlet.

[1] 1 Pa = 10 dynes/cm^2

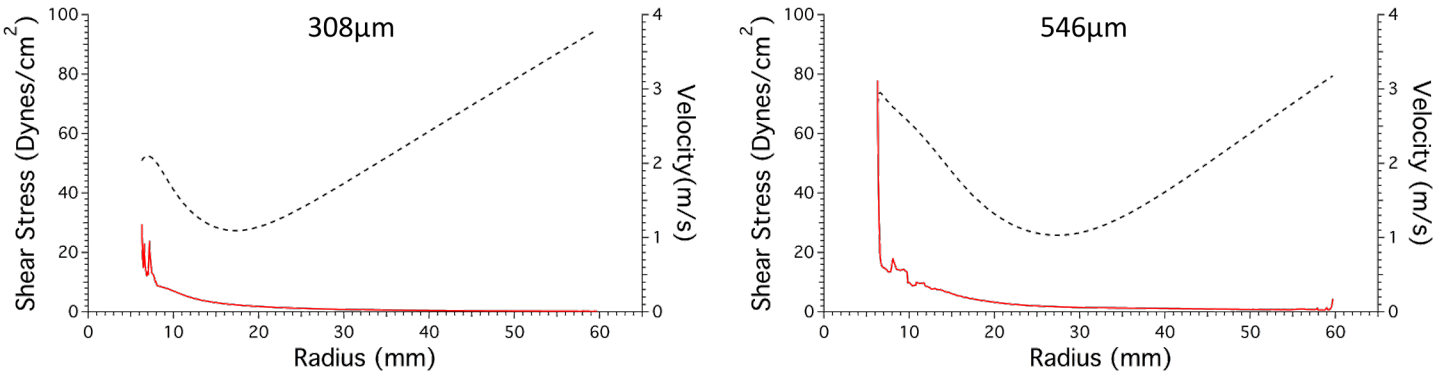


Figure 4.18: Illustration of the change in shear stresses (in red, dynes/cm^2) and velocity (black dotted line, in m/s) as flow progresses through the disc space

Whilst the 546μm achieves the plug-flow profile promised by its design equation to meet an E_K of $\frac{\pi}{2}$, it does so with a spike of high shear stress at the inlet, reaching values seen in small arteries and arterioles (50 – 80 dynes/cm^2). Note also the small rise in shear stress again at the outlet.

The 308μm disc separation calculated to give an N_{Re} of π results in much lower shear stresses at the inlet ($\leq$30 dynes/cm^2), slightly above those of small veins, but not in a smooth fashion. There is no uptick in the shear stress at the outlet.

Figure 4.19 shows a heat map of shear stresses to demonstrate the development of the boundary layers at the inlet. Note that these layers fail to coalesce into a single uniform layer as expected from the design equations.

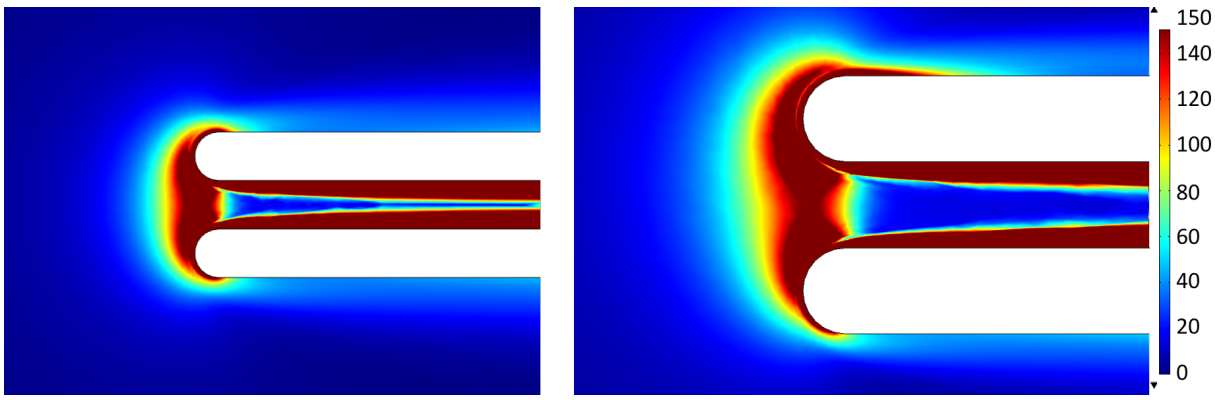


Figure 4.19: Heat map of the shear stresses (in dynes/cm^2) at the disc inlets for a pressure head rise of 120mmHg, demonstrating the boundary layers

Figure 4.20 contains a magnified view of the locations of the peak shear stresses at the inlet.

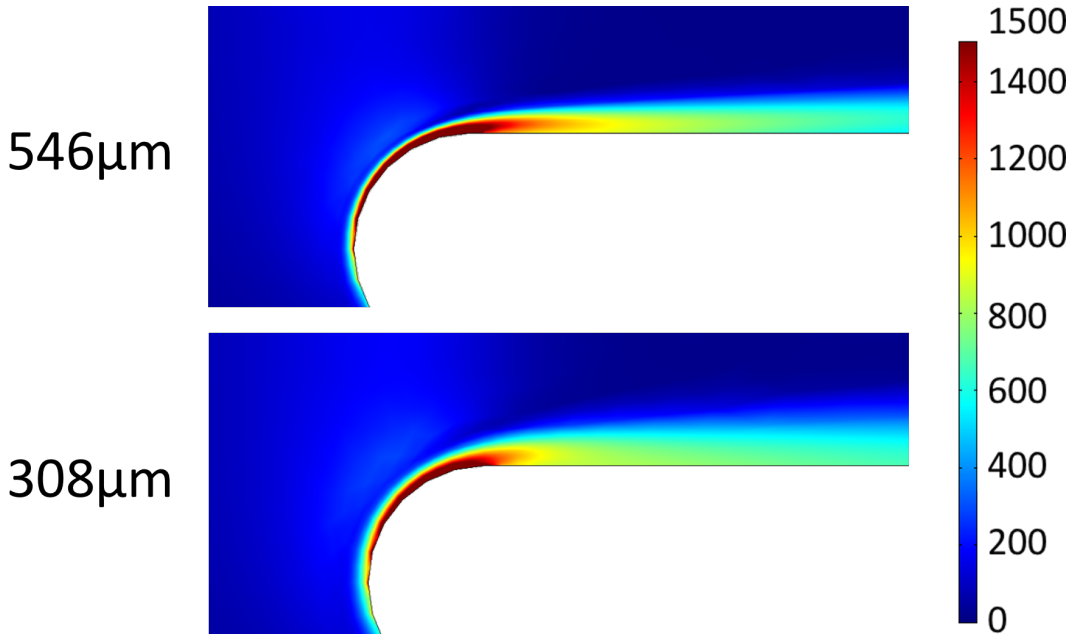


Figure 4.20: Heat map of the peak shear stresses (in dynes/cm^2) at the inlets with 0° inlet angulation, showing a larger area of peak shear stress for 546μm disc separation than 308μm

Note the region of greatest shear stress wrapped around the inlet curvature, with values $\geq$1500 dynes/cm^2 in this small region. However, the size of the peak shear stress region is smaller with 308μm separation than 546μm. Recall that shear stresses in small arterioles can reach 600 dynes/cm^2 [88], with hæmolysis uncommon below 1500 dynes/cm^2.

Flow Coefficients

It was also possible to derive an indication of the energy transfer efficiency from the relative velocities of the disc versus the average radial velocity at the outlet in the COMSOL model using Hasinger & Kehrt's 'flow coefficient', ϕ.

For the 308μm spacing, the average radial velocity is 0.174 m/s, versus 0.274 m/s for 546μm.

$$\begin{aligned}\phi_{308\mu\text{m}} &= \frac{0.174 \text{ m/s}}{66.3356 \text{ rad/s} \cdot 0.06 \text{ m}} \\ &= 0.044 \\ \phi_{546\mu\text{m}} &= \frac{0.274 \text{ m/s}}{66.3356 \text{ rad/s} \cdot 0.06 \text{ m}} \\ &= 0.069\end{aligned}$$

Hasinger & Kehrt recommend a ϕ value of 0.02 or less for good efficiencies[106]. It is clear that the δ value derived from Matsch & Rice's N_{Re} provides better efficiency than Breiter & Pohlhausen's equation with an E_k of $\frac{\pi}{2}$.

4.10.4 Optimal Disc Spacing

These simulations have demonstrated that Matsch & Rice's equation for finding the disc spacing based on optimising for N_{Re}[116] appears superior to that of Hasinger & Kehrt[106] who optimised for plug-like flow in the channel after Breiter & Pohlhausen's original study proposed the same[115].

Interestingly, Hasinger & Kehrt make mention of improved efficiencies in their initial prototype when their E_K value was less than $\frac{\pi}{2}$, a finding seemingly evidenced and replicated here through simulation.

4.10.5 Converging on a Final Design

At this stage it was decided to adopt a manufacturable disc separation of 300μm.

Repeating the simulations with a δ of 300μm demonstrated very good correspondence with the velocities and peak shear distributions for the 308μm studies above.

Max inlet velocity	= 2.05m/s
% Velocity decrease	= 47%
Flow coefficient (ϕ)	= 0.042

It is also possible to use a combination of Breiter & Pohlhausen's equation for the*Ekman number* and Oliveira and Páscoa's pressure equation in order to produce a plot of the expected flow profiles across the biological pressure range given this value ofδ, as shown in Figure 4.21 below.

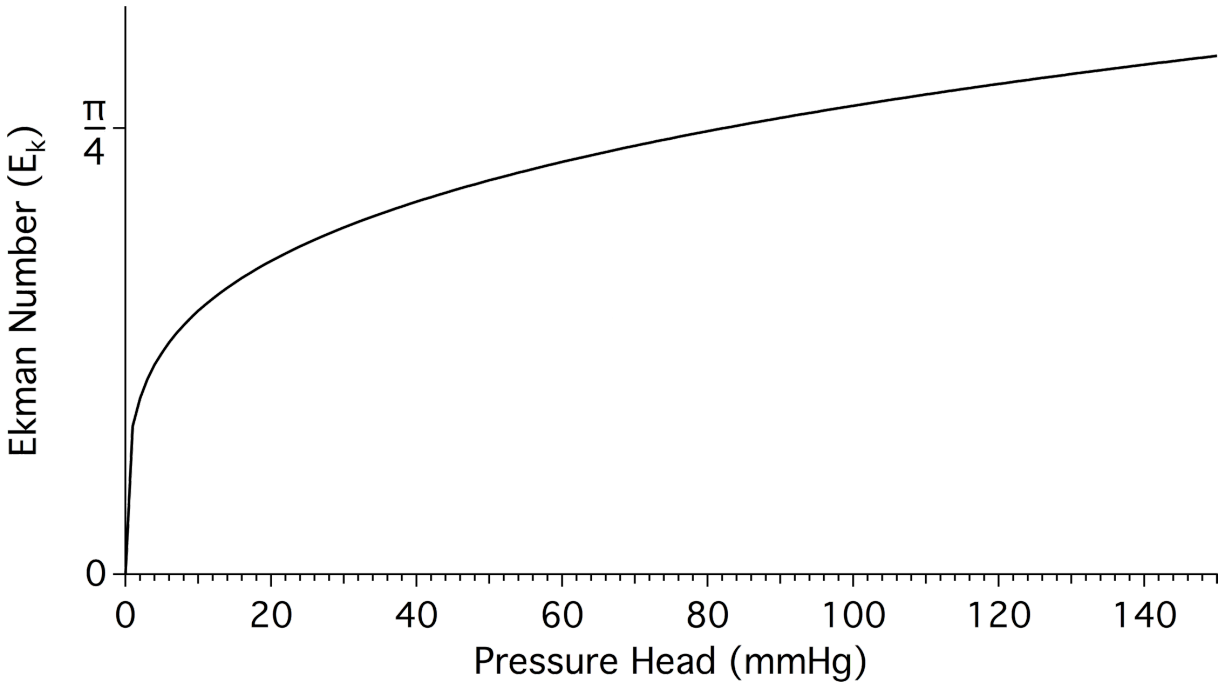


Figure 4.21: Plot of the Ekman Number (which describes the flow profile) vs. the desired pressure head. Refer to Figure 4.7 on page 117 for details

This graph indicates that the given design parameters should give parabolic flow profiles with no reversal across the entire operating range.

4.10.6 Calculating the Required Number of Discs

We can calculate the per-space flow (q) required to meet an A value of 5 according to Hasinger & Kehrt's design charts:

$$
\begin{aligned}
A &= \frac{\delta q}{\nu {r_i}^2} \\
q &= \frac{A \nu {r_i}^2}{\delta} \\
&= \frac{5 \cdot (2.003 \times 10^{-6}\mathrm{m}^2/\mathrm{sec}) \cdot 0.006^2\mathrm{m}^2}{300\mu\mathrm{m}} \\
&= 1.202 \times 10^{-6}\mathrm{m}^3/\mathrm{sec}
\end{aligned}
$$

With a target flow rate of 500ml/minute, we can derive the number of disc spaces (z), and thus the total number of discs as follows

$$Q = q \cdot z$$
$$z = 6.93$$

This value is a non-integer, but close enough to the next nearest integer value (7 spaces), that it is reasonable to accept the calculated flow rate of 504ml/min with 8 discs.

4.10.7 Final Design Data

This design was simulated as before in COMSOL before being prototyped, using the following design data derived from the previous simulations:

μ = 2.033cP
ρ = 1015kg/m^3
r_i = 0.6cm
r_o = 6cm
P_t = 120mmHg (including the 20% *'safety factor'* for pump losses)
ω = 630 rpm (66.33 rad/s)
δ = 300μm
Disc Thickness = 300μm
Disc Number = 8

4.10.8 Disc Stack Simulations

The stack of 8 discs, each measuring 300μm in thickness, and with the same separation were again simulated in COMSOL as a 2D axisymmetric model. Here the disc pump was designed to eject into a larger fluid volume in order to study the flows at each disc space. The inlet was defined as a region of 0 pressure, and the outlet was a large volume bounded by symmetries and distally by a wall. This configuration is shown in Figure 4.22

Note that the bottom disc was solid and without an inlet of radius r_i. This was both for ease of manufacturing, but also simulations with open bottom discs often failed to converge, or had large backflows in the space between the bottom disc and the housing.

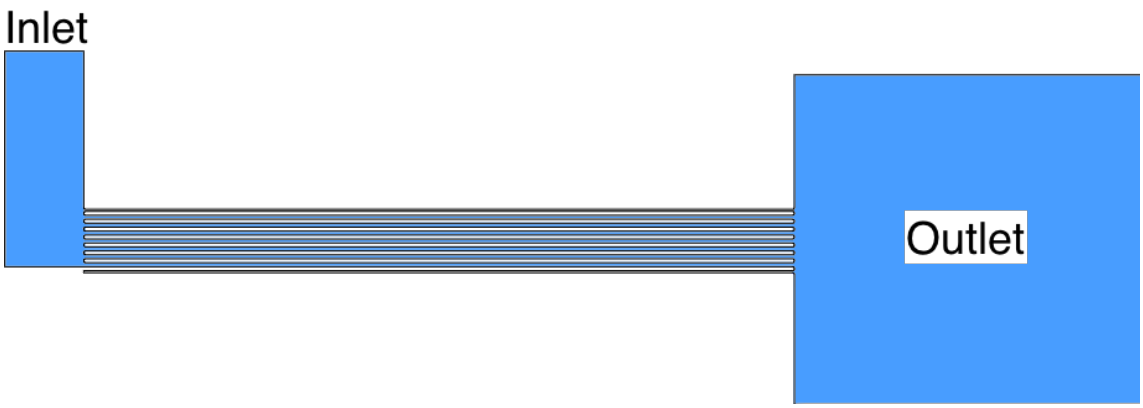


Figure 4.22: Final model for simulation of the 8-disc stack in COMSOL showing the solid bottom disc, an inlet of radius r_i and the larger volume of the outlet

Outflow Convergence

The first thing to note was that the outflow from the pump converged into a single stream emerging from the midline of the disc stack, undoubtedly a result of the boundary layers coalescing from each disc space, shown here in Figure 4.23.

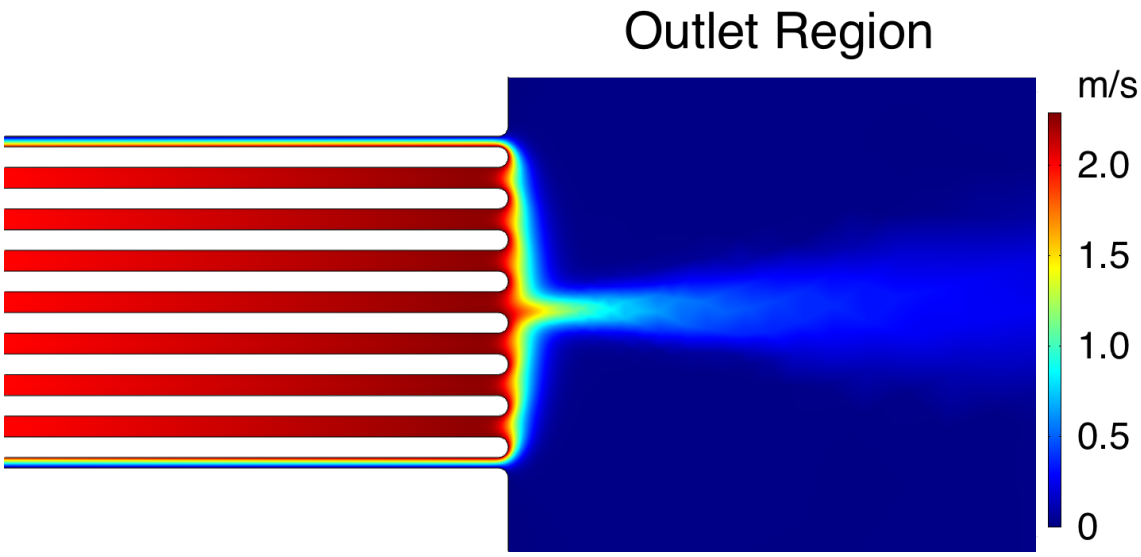


Figure 4.23: Outflow of the Tesla pump coalescing into a single jet emerging from the centre of the disc stack

This effect also appears in the flow lines from Izraelev *et al.*'s Tesla pump LVAD, shown in Figure 4.6 on page 114, indicating that this may be a general feature of Tesla-type pumps.

Radial Flow Profiles

The radial flow profiles (exaggerated 5× in Figure 4.24 for clarity) were not as predicted by the previous simulations of single disc spaces, nor by the Ekman numbers which should have described non-reversed, parabolic profiles. The profiles are curvilinear and a combination of the radial flow between the discs and a 'wicking' effect of the boundary layer coalescence at the outer radius. Note also the pronounced Couette flow profiles between the stationary housing and the upper and lower discs.

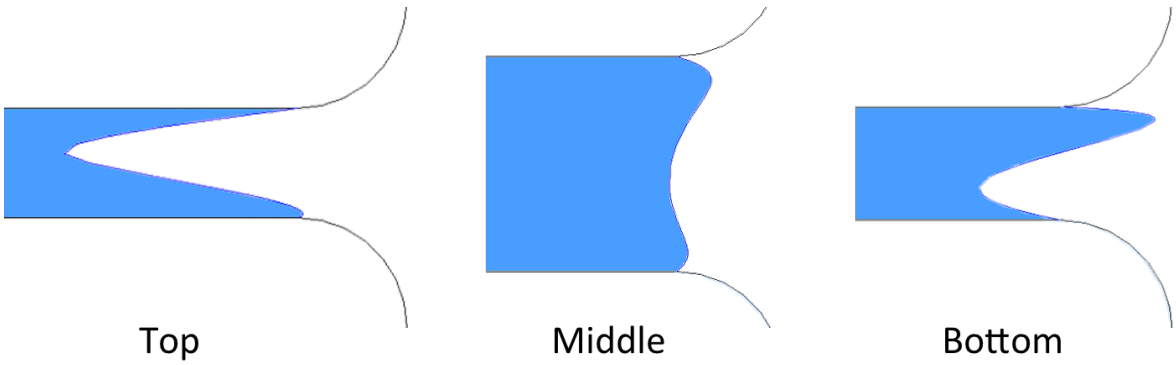


Figure 4.24: Selected radial outflow profiles from the disc stack in blue, exaggerated 5×

There was also a pronounced difference in the outflow velocities between each disc space as shown in Figure 4.25, with the central space providing the lowest outflow compared to the upper and lowermost spaces. This may be again due to a 'wicking' effect of the coalescing boundary layers drawing more volume through the upper and lowermost spaces into the developing outflow jet.

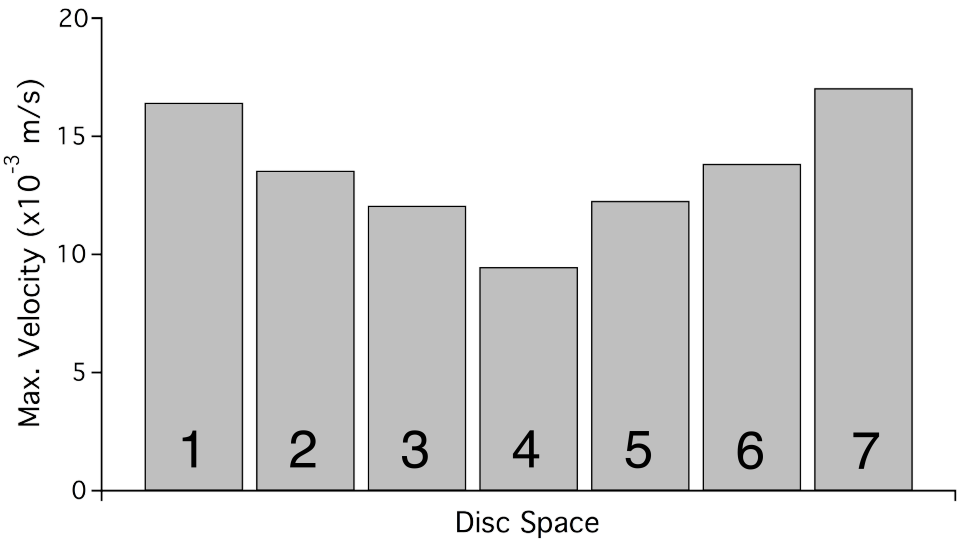


Figure 4.25: Per-space variation in disc outflow velocity. Disc space 1 is the lowermost

Couette Backflow

Finally, despite minimising the spacing between the moving and immobile surfaces, it appears that any future designs will also have to incorporate a form of baffle to prevent recirculation of the Couette backflow shown in Figure 4.26.

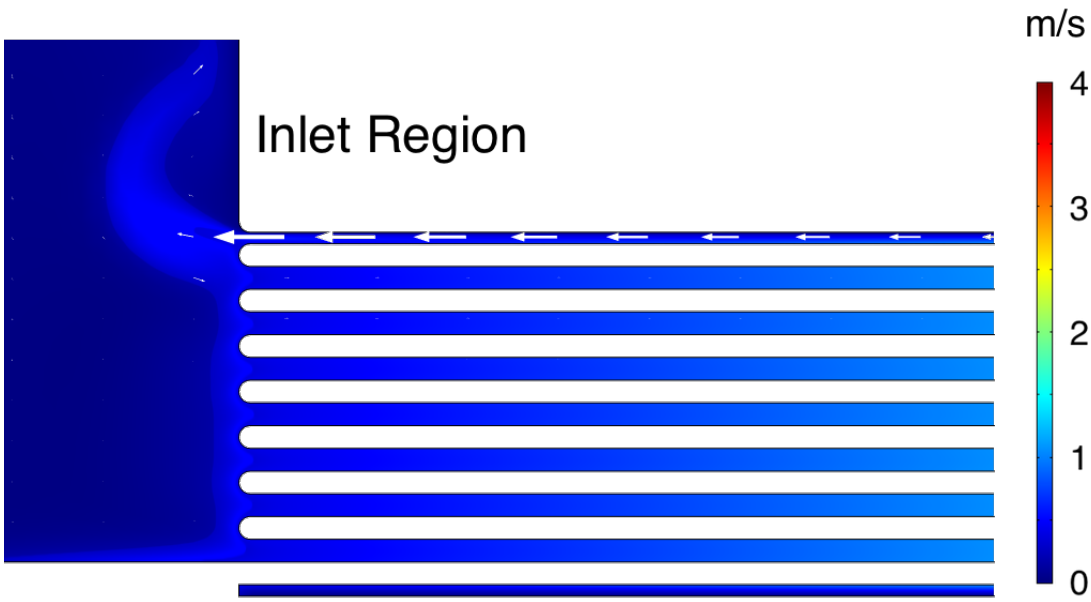


Figure 4.26: Highlighting the back-flow through the Couette region between the uppermost disc and the stationary housing, with a tongue of liquid flowing back up towards the inlet. Velocity scale in m/s

These secondary flows appear to be accommodated in other designs such as those of Izraelev *et al.*[108] and the later revision to their work by Jhun *et al.*[114]. However, at full flow in this simple Tesla pump, the backflow jet may still interfere with flow through the uppermost disc space.

4.11 3D Volute Simulations

Even with this improved design there was a concern that a simple 3D volute design would be unable to properly direct the flow from the disc stack and adequately convert this to pressure at the outlet.

Therefore, help was sought through a collaboration with Professor Yun Xu in the Department of Chemical Engineering, and her post-Doctoral Research Associate Dr. Qunfeng Zhang. Their experience in designing turbomachinery proved very helpful in exploring the simulation in 3D. I was able to guide the simulations in order to identify some key problems in 3D volute design for a Tesla-type pump.

4.11.1 Simulation Results & Conclusions

Dr. Zhang began by transferring my model from SolidWorks into STAR-CCM+®v10.02[1], a computational fluid dynamics package more suited to 3D simulations of pump geometries than COMSOL. The pump was simulated at its peak expected output pressure of 150mmHg.

These simulations demonstrated that the pump was operating as expected with respect to sub-hæmolysis shear stresses along all disc surfaces, demonstrated in Figure 4.27 below.

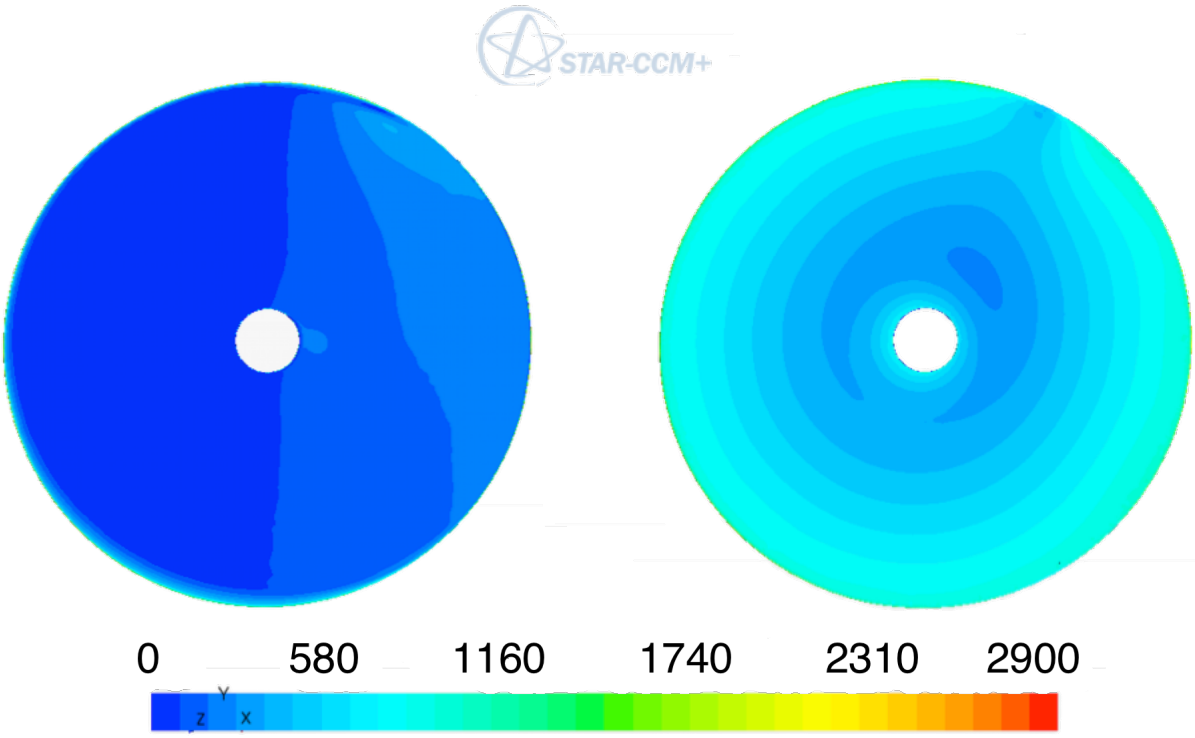


Figure 4.27: Shear stress on the lower (*left*) and upper (*right*) surfaces of the top disc in the stack, with contours pointing into the region of the outlet. Scale in dynes/cm^2

On each disc, the shear lines appeared to point and gather towards the pump outlet, and the shear stress magnitude was certainly many times greater on the surface of the uppermost disc as a result of the Couette flow between the upper surface of the top disc and the stationary housing.

An important realisation came with the simulation of flow through the pump, shown as a projected image in Figure 4.28:

[1]CD-adapco, Melville, New York, USA

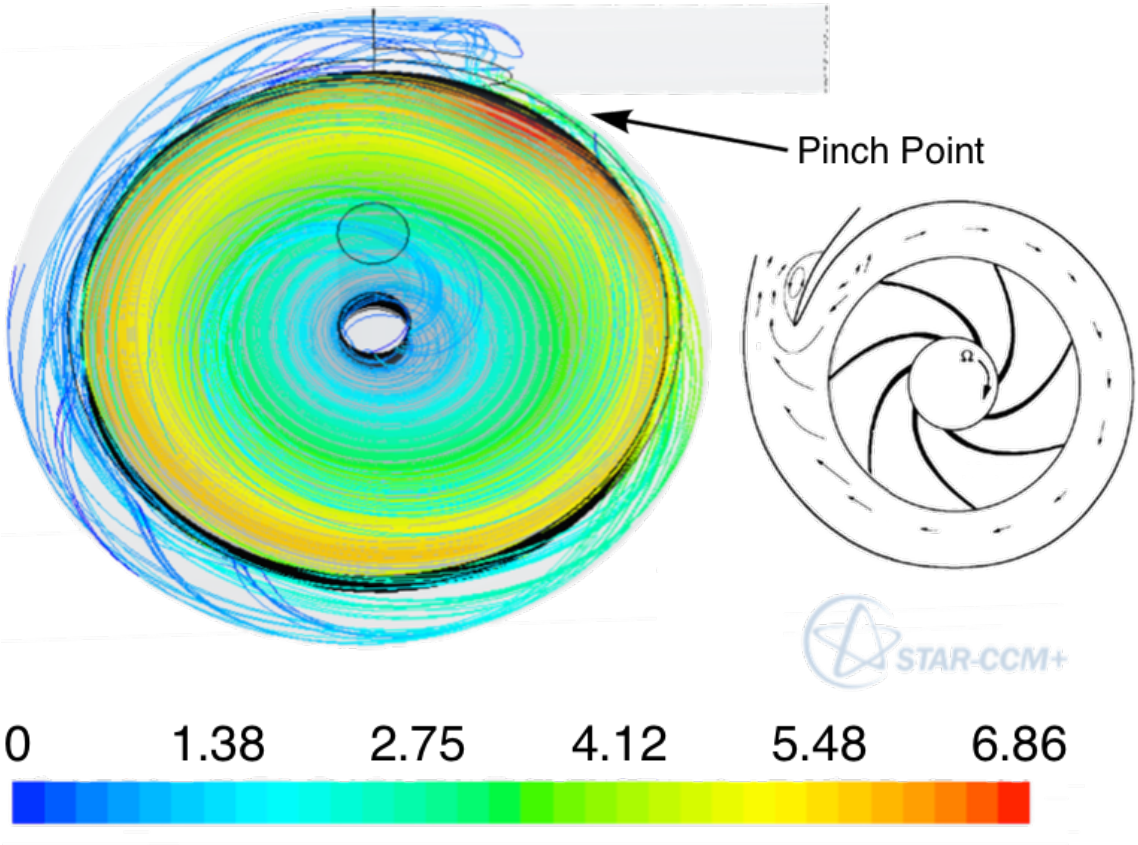


Figure 4.28: Velocity flow lines in the 3D simulation of the volute indicating a 'pinch point' at the tongue (arrow), and multiple trajectories recurving back into the region of the inlet. Values in m/s. *Inset*: illustration of abnormal flows around the tongue from Figure 4.5 on page 111

This study demonstrated that the liquid is initially drawn from the inlet and enters into the inter-disc spaces as per design. It circulates around the disc multiple times gaining energy, until it reaches the disc periphery at a peak of 5.5 – 6.8 m/s where it enters the volute. Here the energy is rapidly lost within a quarter of a full rotation, never traversing the full volute to the outlet and instead stagnating and recirculating where able, including through the upper disc–housing space.

This can be seen by tracing the trajectory of an individual flow pathway from the periphery of the disc to where it rapidly loses velocity, turning blue, before recurving into the region of the disc stack inlet.

Not only does the liquid fail to flow through the pump effectively with a standard spiral volute design, it also leads to increased shear stresses at a 'pinch point' just beyond the outlet, reflected as the focal increase in the peak velocity in Figure 4.28, and the pointing of the shear stress maps from the surfaces of each disc in the stack (Figure 4.27).

One final simulation involved unrolling the scroll by a further 5% so as to open out the pinch point beyond the outlet. Unfortunately, this did not affect the recirculation or energy

loss in the volute, as shown in Figure 4.29 below.

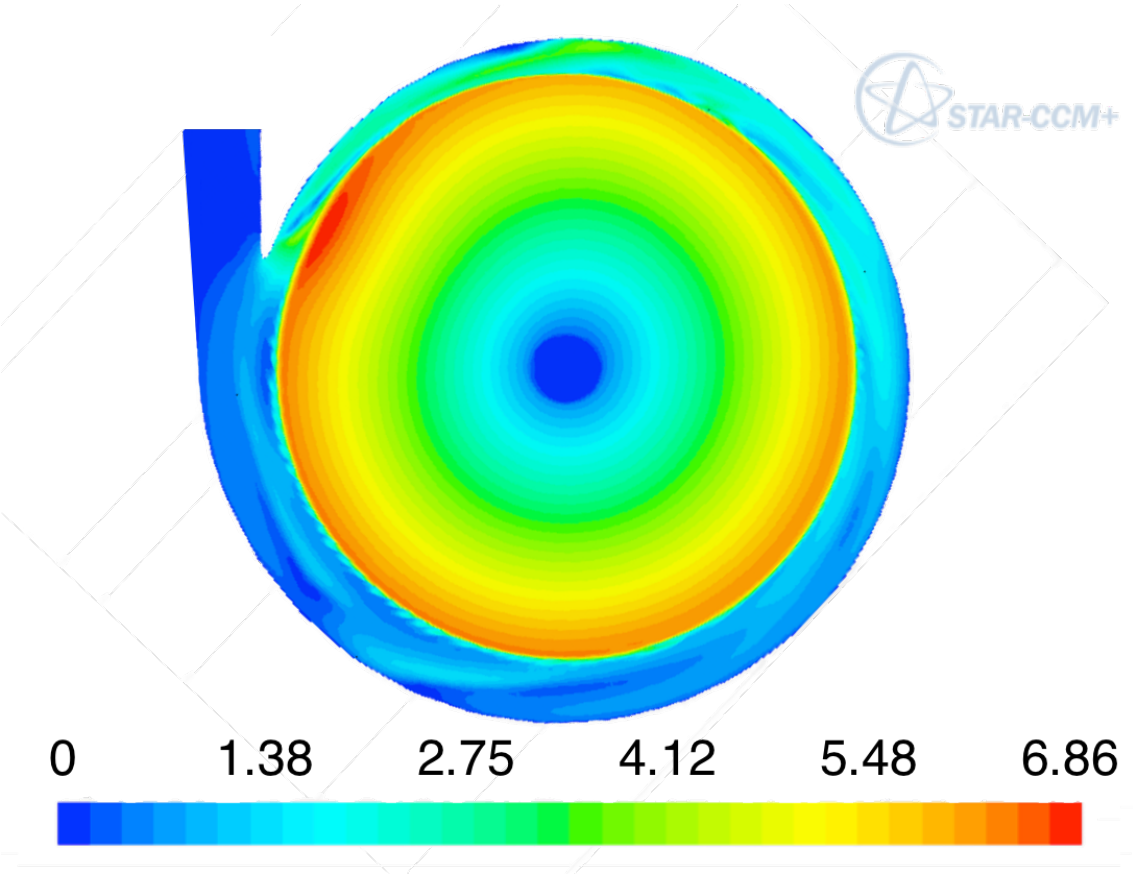


Figure 4.29: Velocity flow lines in the simulation of the wider volute spiral. Values in m/s

The result of both of these volute models (Figures 4.28 and 4.29) was that the pump was able to develop pressure within the volute but no simulated through-flow, an effect also seen in the first physical prototype (Figure 4.13 on page 131).

4.11.2 Conclusions from 3D Simulations

It appears that the housing design principles that apply to standard bladed impeller-based centrifugal pumps may not apply to multi-disc Tesla-type pumps, and that the problems of head loss through turbulence, flow recirculation, and Couette backflow are not as straightforward as initially imagined.

As described on page 135, calculations of the Reynolds numbers for the flow in the region just beyond the tips of the spinning discs describes turbulent flow regime, with moderate intensity. The fluid will therefore lose energy during the transition from the laminar flow within the stack, to a turbulent flow regime upon exit.

Interestingly, the laminar flow from the interspaces appears to coalesce into a single

outflow jet from the mid-plane of the disc stack. However, as can be seen in the designs of Izraelev *et al.*'s in Figure 4.6 on page 114, any vertical dead space in this region appears to encourage regions of stagnation and recirculation, further promoting turbulence and undoubtedly increasing the magnitude of any head loss.

Future volute designs may therefore need to guide and baffle the laminar outflow from the disc stack so as to minimise this region of turbulence and energy loss. Izraelev*et al.* have already given an indication of some of the benefits that can be achieved by contouring the outer radius of the disc stack to equalise the flow from each interspace, yet their final volute design remains far from perfect and still demonstrates regions of recirculation and stagnation.

Finally, the differential between the low pressure of the inlet region of the disc stack and the higher pressure beyond the tips of the spinning discs (despite the energy losses through turbulence and recirculation) leads to the development of Couette backflow between the uppermost disc and the stationary housing. This was first encountered in my initial design and discussed in more detail on page 131.

These secondary flows appear severely detrimental to the through-flow of a Tesla pump, and increasing the angular velocity of the disc stack to try to compensate for any head loss would only exacerbate the magnitude of these adverse flows by increasing the reverse pressure gradient. Perhaps a future design can be realised wherein the housing itself spins with the disc packet, or internal baffles are used to otherwise prevent these backflows.

Pumps vs. Turbines

This is one key area where Tesla turbine and pump designs appear to diverge – slip-flow through a turbine simply leads to a loss of efficiency, whereas un-entrained and adverse flows in a blood pump configuration lead to recirculation, head loss, and potential hæmolysis. This is an area that has not previously been addressed in the literature and may lead to an emergence of different design parameters for each configuration.

Focus was then returned to attempting to optimise flows at the inlet to the disc stack instead.

4.12 Exploring Changes to the Inlet Angle

It is reasonable to expect that it would be possible to minimise the energy loss at the inlet by adjusting the angulation of the discs from the inlet to the 'neck', until the tangential

flow profile had developed sufficiently to be the major contributor to the total velocity of the liquid.

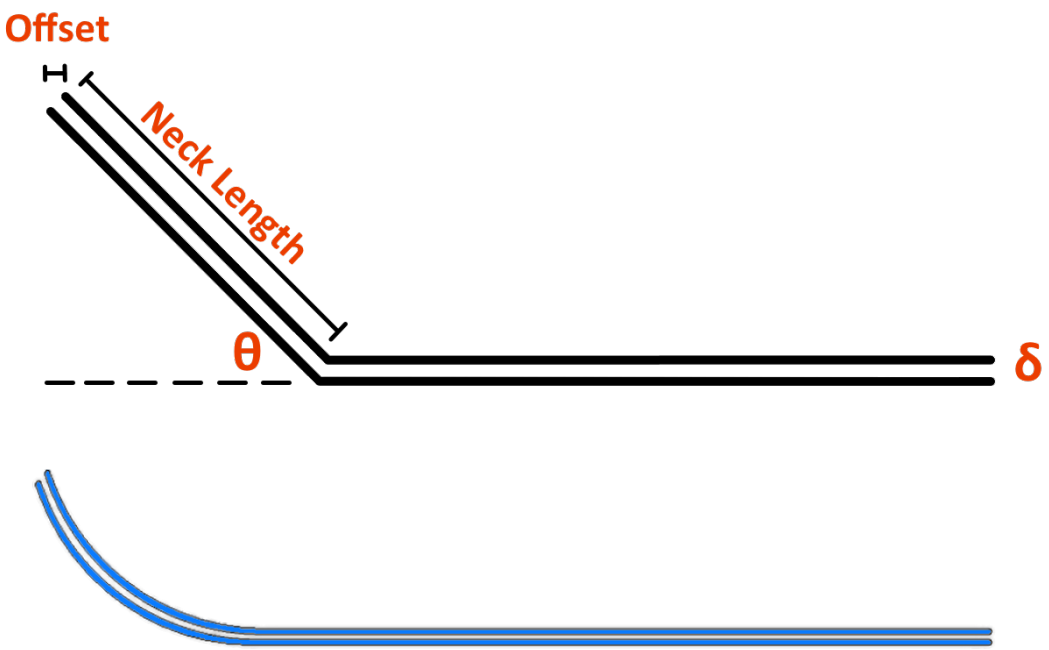


Figure 4.30: Definitions for the variations in the disc inlet angle and an example of discs with 308μm separation, δ inlet offset, and 45° of angulation

Studies were undertaken for increasing angles of inclination of the inlet from the plane of the disc (θ), at values of 0°, 11.25°, 22.5°, 45° and 60°. The radial offset of the bottom disc vs. the upper disc were also varied from δ to 2δ, with the inter-disc spacing then converging to the standard δ disc separation.

COMSOL's internal `Geometry > Fillet` function was used to produce smooth continuous curvatures from the inlet point defined by the angulation and length of the neck segment to the flat portion of the discs.

Boundary Layer Separation

The first series of simulations explored the relationship between the structure of the boundary layer at the inlet and the increasing angulation of the neck. It became clear that the boundary layer separated from the underside of the upper disc at *all* degrees of angulation with 546μm of disc separation (δ) and any amount of offset, and for all degrees of angulation with 308μm of disc separation when there was 2δ radial offset.

However, with 308μm of disc separation and δ radial offset at the inlet, there was *no* boundary layer separation from the underside of the upper disc. The images in Figure 4.31

demonstrate minor and severe boundary layer separation with increasing disc spacing, inlet radial offset and neck angulation:

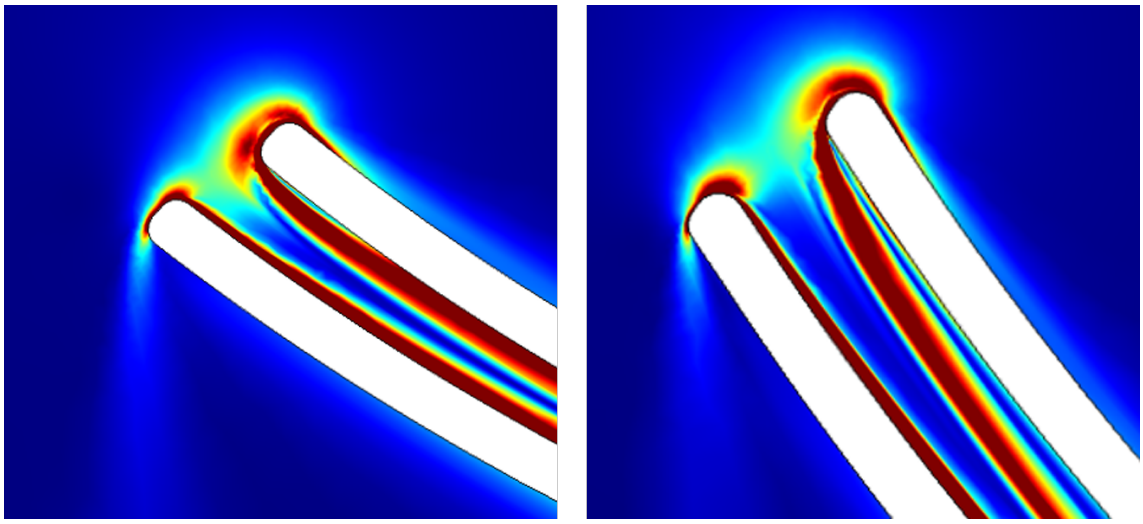


Figure 4.31: Demonstration of boundary layer separation. *Left:* 308μm separation and 616μm offset, 22.5° angulation – slight separation before rejoining the disc. *Right:* 546μm separation, 1092μm offset, 45° angulation – total separation

Improved Energy Transfer

Under ideal conditions, there would be no energy loss at the disc inlet and the velocity of the flow would rise monotonically from the inlet to the outlet, demonstrating continuous energy transfer. Angling the inlets should allow gravity to impart additional energy to the incoming stream, maintaining the radial velocity until the liquid begins to move tangentially with the boundary layer.

This was investigated for the configuration that led to zero flow separation at all angles - 308μm disc separation and inlet radial offset, at increasing inlet angulation.

The graph in Figure 4.32 illustrates the change in the total velocity along the midline of the 308μm channel, passing from the inlet to the outlet. Due to the curvature of the geometry, the total radial path length exceeds 54mm (60mm r_o – 6mm r_i) as θ increases.

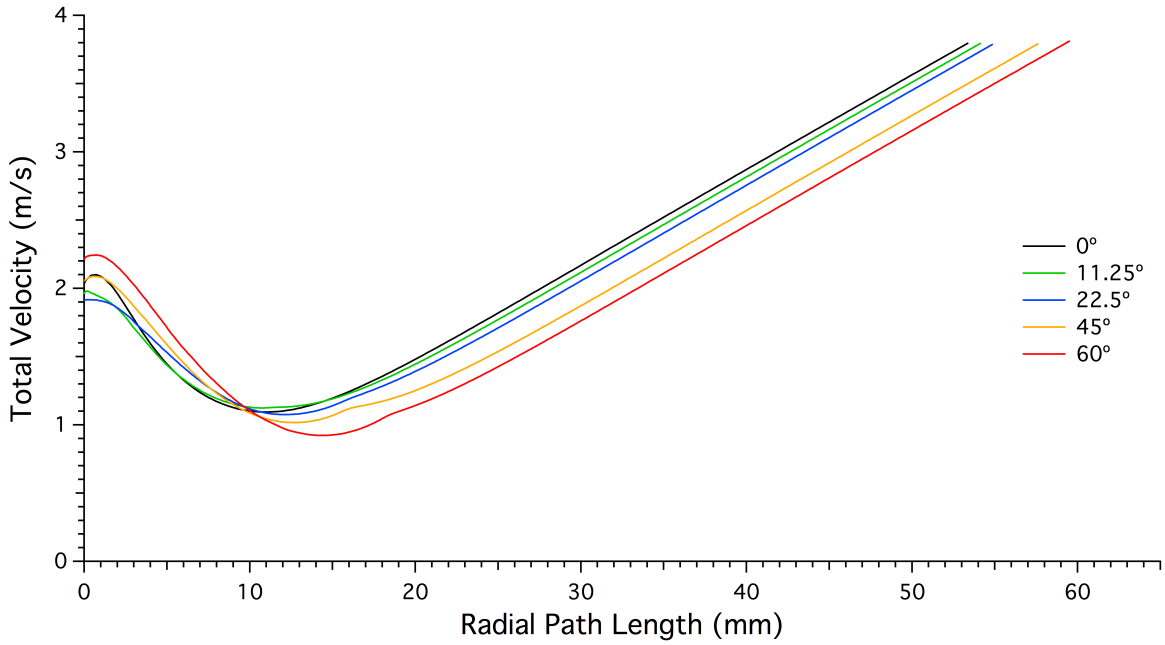


Figure 4.32: Total flow velocity vs. increasing disc inlet angle (θ), with a disc separation of 308μm and the lower disc offset by δ

There appears to be an initial decrease in the inlet velocity from 0° – 22.5° angulation, rising again at 45° & 60° of angulation.

Table 4.2 below compares the magnitudes of peak inlet velocity, the percentage decrease at the nadir, and the peak shear stress in the simulation field for 308μm vs. 546μm disc spacings with an offset of δ.

Table 4.2: Flow and Shear Stress Parameters for 308μm vs. 546μm Disc Separations

	Inlet Velocity (m/s)		% Decrease		Peak Shear Stress (Dynes/cm^2)	
Angle	308μm	546μm	308μm	546μm	308μm	546μm
0°	2.1	2.95	48	65	3332	5137
11.25°	1.98	3.03	52	67	2221	3365
22.5°	1.73	2.73	44	68	2177	3531
45°	2.09	2.87	51	67	3709	4241
60°	2.25	3.07	58	72	5528	7761

It is clear that a disc separation of 308μm is superior to that of 546μm, particularly with regards to the energy lost from the incoming flow represented as the % decrease in velocity.

This disc design also gives a 'flow coefficient' (ϕ) of 0.049 which is not quite as good as that for a flat pair of discs (0.044), but the design does bring additional benefits.

4.12.1 Optimal Disc Design

It appears that an optimal disc design should also include slight of angulation of the inlet, in the amount of 22.5°, so as to minimise the total shear stress in the field as well as the energy loss at the inlet.

The final pair of images in Figure 4.33 demonstrates the smaller area experiencing peak shear stress with a 22.5° angled inlet versus a flat disc stack:

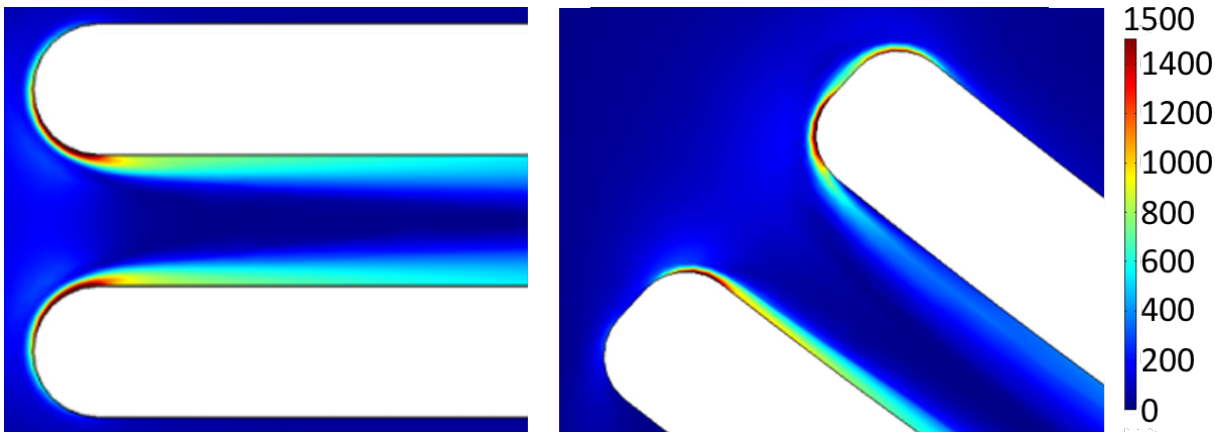


Figure 4.33: Reduction in area of peak shear stress at the inlet by 22.5° angulation (right), with a disc separation of 308μm and the lower disc offset by δ. Values in dynes/cm^2

4.13 Concluding Remarks and Future Work

Tesla-type pumps have only been investigated for use as blood pumps by a small number of groups to date[107, 108, 112, 113] but the underlying boundary-layer viscous drag mechanism is theoretically promising for a shear-thinning liquid like blood.

It also appears that the stacked-cone design of the BPX-80 Bio-Pump® Plus from Medtronic employs a rudimentary form of Tesla's principles, using boundary layer effects to pump the liquid.

These studies have demonstrated the following:

1. Optimising the inter-disc spacing according to Matsch & Rice's requirement for the optimal *operating Reynolds number* (N_{Re}) to lie between 2 – 4 provides better boundary layer adhesion and lower shear stresses than Breiter & Pohlhausen's design for an E_K of $\frac{\pi}{2}$
2. A slight degree of angulation of the inlets (22.5° in this study) is desirable, whilst maintaining a uniform curvature and disc separation. Excessive curvatures and disc offsets lead to boundary layer separation from the underside of the discs

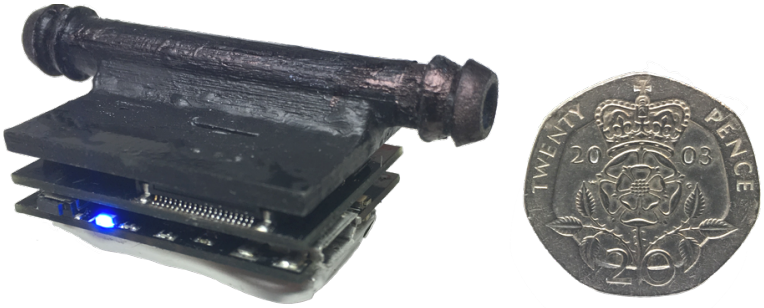
3. Boundary layer effects cause liquid outflows to coalesce and 'wick' from the outer radius of the discs, converging into a single jet flow. This could strongly influence future design decisions for the outflow region of multi-disc pumps
4. Traditional scroll volutes do not adequately capture or transfer the energy gained by the liquid in the disc stack into pressure and flow at the outlet
5. Tesla-type pumps are severely impacted by entrapment of air at the inner radius, leading to stalling and turbulent recirculation
6. Secondary Couette flows between the moving discs and the immobile housing lead to back flow and recirculation

Building a better Tesla pump from these findings would involve closer attention to the volute, perhaps with an 'inverse turbine' or 'rose petals' arrangement of blades angled to capture outflow from the stack at multiple points simultaneously around the circumference before converging on a single outlet in the same axis as the inlet.

Magnetolevitation to remove bearings and potential sources of leakage from the stack would also seem a reasonable optimisation for future embodiments of the design.

Finally, a formal mathematical account of the Couette flows between the moving discs and immobile housing and their contribution to the overall design of the pump is an area that has not been tackled by any author publishing in this field to date and would seem ready for exploration.

Part 2:
The Multimodal SmartPipe Device



Chapter 5

Introducing the SmartPipe

This Chapter introduces the concept of the SmartPipe and how it fits into the concept of normothermic machine perfusion of kidneys for transplantation.

Subsequent Chapters dissect the device and describe the design and implementation of its various internal components, including the custom software and user interface.

5.1 Organisation of this Part of the Thesis

The Chapters in this part of the Thesis dissect the SmartPipe device layer-by-layer, describing the relevant design and implementation of each component to achieve the goal of a small, portable multimodal physiological sensor suite for blood perfusion circuits.

Chapter 5 Introducing the SmartPipe concept, the parameters being monitored, and the design process

Chapter 6 Selection of the data collection system and designing a custom DAC to interface between the SmartPipe and this system

Chapter 7 Designing the bottom (digital) board including the power subsystem, the software and the user interface

Chapter 8 Selecting the sensors and the design of the interface pipe for the top (analog) board

Chapter 9 Designing the middle (mixed-signal) board for translating the analog sensor responses into digital data, results of the sensor tests, and concluding remarks

5.2 Monitoring Perfusion Circuits

Adapting established cardiopulmonary bypass circuits into miniaturised, low-cost, storage and 'nursing' devices for the normothermic perfusion of kidneys destined for transplantation is not entirely straightforward. There is a distinct need to be able to sense and monitor a collection of general physiological parameters with a low-cost, miniature sensor suite to ensure that it is successfully mimicking the native environment of the kidney[123], and to be able to implement feedback control loops should any such parameters drift out of range.

There are systems on the market for continuous multi-parametric monitoring of cardiopulmonary bypass circuits, such as the CDI™ Blood Parameter Monitoring System 500 from Terumo[1]. This device uses an optical blood oxygen saturation probe and a series of fluorescence sensors to continuously monitor 11 different blood parameters. According to Terumo's own marketing information, it is now used in a third of all cardiac surgery cases in the US.

However, the CDI device is tailored for use in cardiopulmonary bypass circuits, and whilst the sensor head is small, it requires an external 7.3kg processor and monitor, as well as a 3.8kg calibration unit and specialised gas bottles. The second-hand purchase price for the

[1]Terumo Corporation, Shibuya, Tokyo, Japan

monitor alone is approximately US$6,000[1].

5.3 The SmartPipe Concept

This part of the project therefore explored the possibility of collecting and analysing live temperature, pressure, flow and blood oxygenation levels in a perfusion circuit via a low-cost, small, battery-powered multimodal sensor module built into the tubing itself – the SmartPipe. An overview of this concept is shown in Figure 5.1 below:

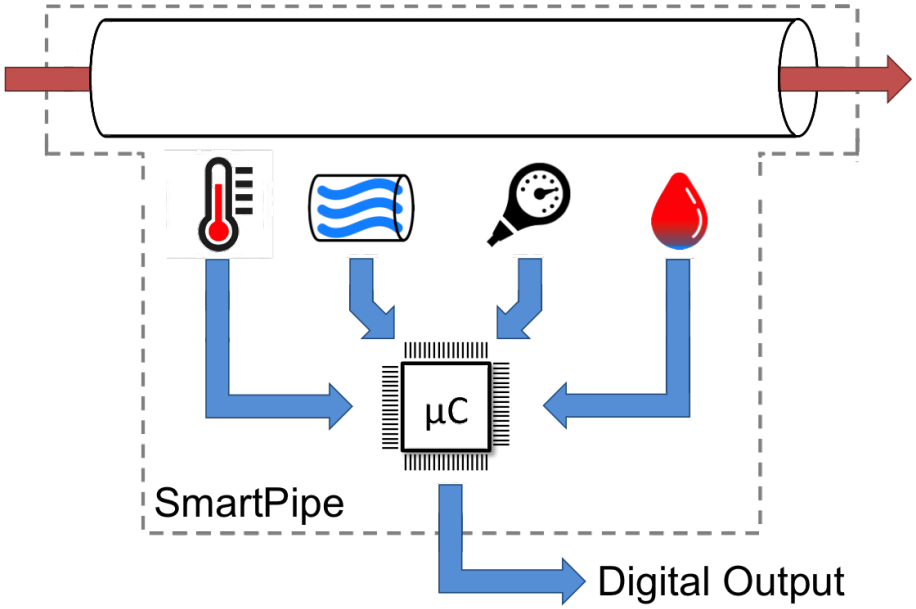


Figure 5.1: High-level overview of the SmartPipe monitoring and data collection system demonstrating its ability to collect temperature, flow, pressure and oxygenation data from an erythrocyte-based perfusate

Incorporating the sensors into the wall in this way should avoid any disturbance to the laminar flow and prevent thrombosis or hæmolysis.

Figure 5.2 below demonstrates the imagined use of the SmartPipe in a kidney perfusion circuit:

[1]Priced from units for sale on an internet auction site in December 2015

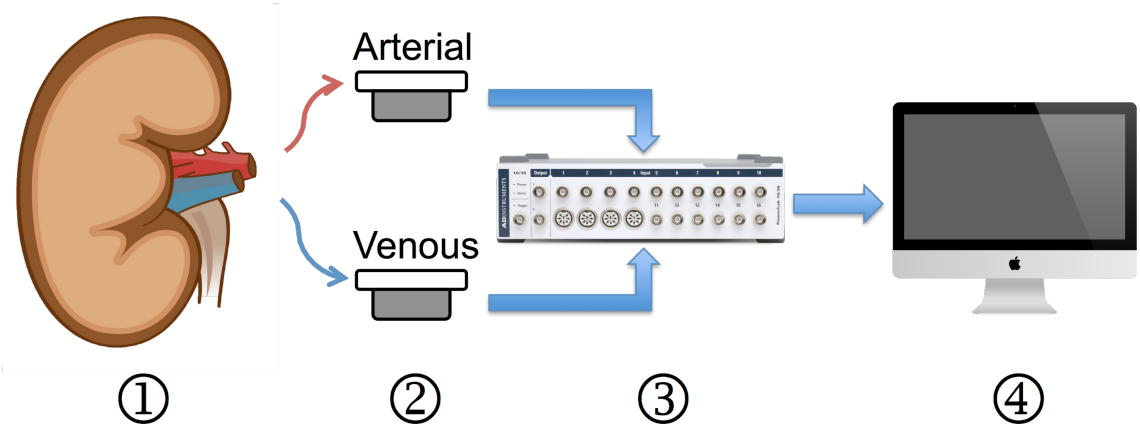


Figure 5.2: The imagined use of the SmartPipe monitoring and data collection system in an extracorporeal perfusion circuit for renal grafts

This system consists of:

1. A kidney undergoing warm machine perfusion
2. Two SmartPipe devices monitoring physiological parameters in the arterial and venous flow
3. Data passing from the SmartPipe devices to a data acquisition system[1]
4. Data displayed and stored on a PC for later analysis

This device could be placed at multiple points around the circuit for both research purposes and for implementing control loops to maintain homeostasis in a closed-loop system.

5.4 Inside the SmartPipe

The final research version of the SmartPipe is a complex device comprising multiple analog sensors exposed directly to the perfusate, specialised detection circuitry, and a microcontroller-driven digital control system for collecting and relaying this data for analysis whilst controlling the power and user interface subsystems.

The device is implemented as a stack of 3 interconnected multi-layered PCBs and a 3D-printed tailored pipe. This minimises the size of the final device and allows the user to plumb it directly into a perfusion circuit. The internal detail of the electronics stack is shown in Figure 5.3 below, with a brief guide to the functionality of each layer:

[1] A PowerLab from ADInstruments of Dunedin, New Zealand

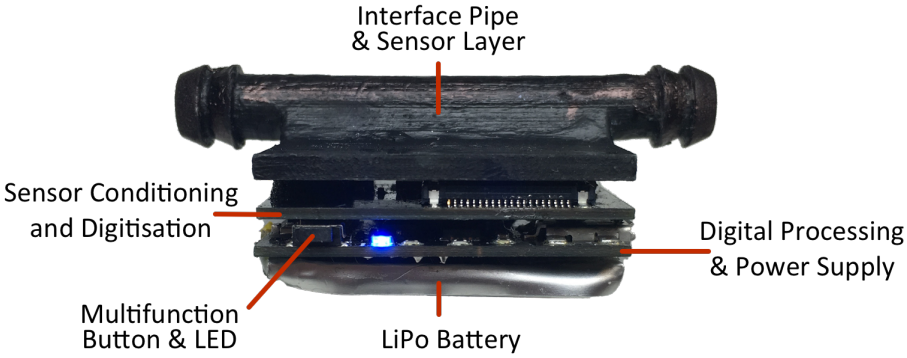


Figure 5.3: Profile view of the SmartPipe demonstrating its internal design with 3 stacked multi-layer PCBs, the disposable interface pipe containing the sensors, the multifunction user interface button, a small lithium polymer (LiPo) battery on the underside, and a micro-USB port on the right of the bottom layer for recharging

5.5 Selecting the Physiological Parameters to Monitor

This section describes the parameters to be monitored by the SmartPipe, the reasoning behind their selection and discusses the physical properties of the signals including ranges, sensitivity and bandwidth to guide the later stages of the design process.

5.5.1 Temperature

Our enzymatic and metabolic processes have evolved to function optimally within a range of internal body temperatures between 36.5°C – 37.5°C. Any organ that has been removed from this environment into a perfusion circuit will need thermal support from a heat exchanger to prevent excessive heat loss to the environment, or overheating and irreversibly denaturing essential proteins at temperatures above 40°C. This heat exchanger should also be able to operate as a cooling device at the research stage to allow comparisons against the existing preservation paradigm.

Maintaining the temperature to within ±0.5°C of 37°C with a control loop therefore needs a sensor capable of detecting changes with 0.1°C resolution in order to provide sufficient steps within this range. The sensor should be capable of detecting temperatures across the range of 0°C – 40°C to provide sufficient sensing overhead for cooling protocol targets of 4–5°C, and the normothermic target of 37°C.

Fortunately, the sensor does not need to operate with high bandwidth as physiological temperature changes occur over minutes to hours, and a good-quality heat exchanger built into the circuit should maintain the temperature of the perfusate within a narrow range with only mild perturbations.

A temperature sensor with 0.1°C sensitivity should therefore be sufficient to monitor anticipated temperature changes of this magnitude, and require a bandwidth of $\leq$1 Hz.

Final Design Targets

Range: 0 – 40°C Sensitivity: ±0.1°C Bandwidth: 1 Hz

5.5.2 Pressure

An isolated normothermic perfusion circuit for a single kidney must be able to replicate the normal perfusion pressures seen by the organ within the body.

As such, the pressure sensor system designed into the SmartPipe must be able to detect changes and magnitudes consistent with normal human blood pressure, and with a bandwidth sufficient to capture significant changes such as a pulsatile pump simulating the heart's systolic ejection stage.

A normal human arterial blood pressure waveform is reproduced here:

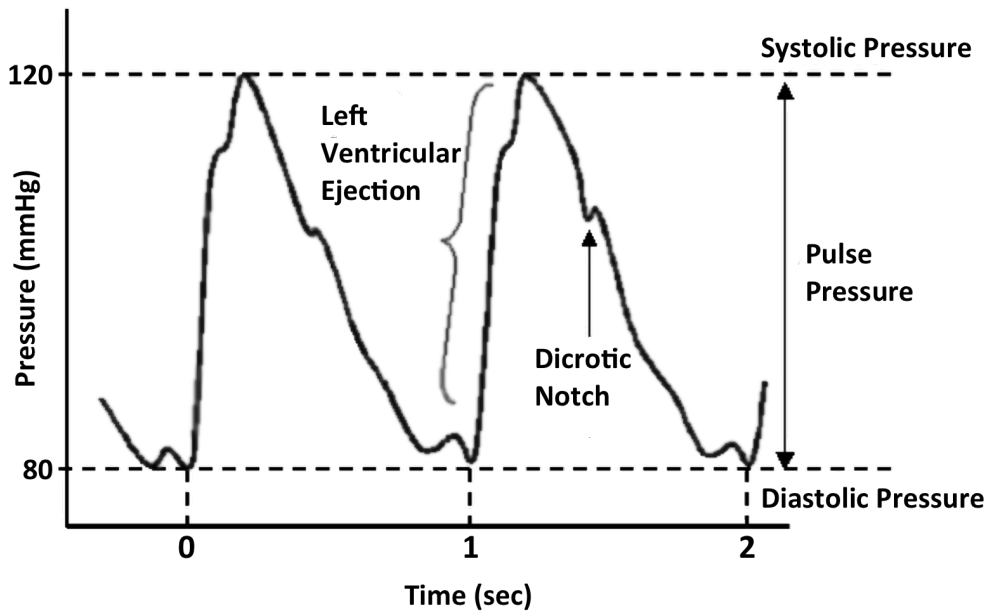


Figure 5.4: A normal human arterial pressure waveform demonstrating the range of systolic and diastolic pressures and the slope of the systolic ejection phase. Note the *dicrotic notch* which indicates aortic valve closure and the beginning of diastole, or cardiac relaxation. From [124]

Therefore, capturing the entire range of normal blood pressures requires a sensor capable of operating between 0mmHg – 180mmHg to provide 50% overhead to allow for kidneys

from hypertensive organ donors, and over-pressures required to overcome microvascular sludging and vasospasm sometimes seen in transplant organs.

With a normal heart rate of 60-80bpm, a pressure sensor requires a bandwidth on the order of 200Hz to operate beneath the Nyquist limit. Established clinical practice allows pressure to be determined to the nearest 2mmHg, and therefore this is the sensitivity that the pressure monitor should aim to achieve[124], although many pressure sensors can achieve better resolutions than this.

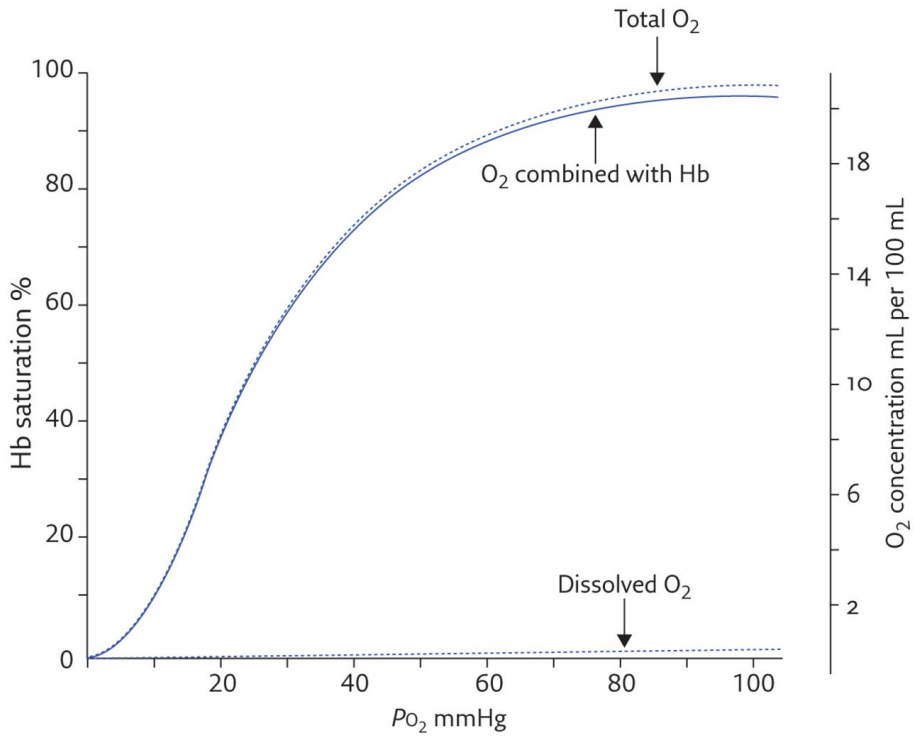
Final Design Targets

Range: 0 – 180 mmHg Sensitivity: 2 mmHg Bandwidth: 200 Hz

5.5.3 Oxygen

Section 2.5 on page 58 discussed the background of renal oxygen metabolism in detail. The goal of the oxygen sensor is to indicate the health of the tissue according to its oxygen consumption rate, a practice which has been adopted by other groups working in the area of machine perfusion[58].

The majority of the oxygen is delivered by erythrocytes which are packed full of hæmoglobin, a specialised carrier molecule comprised of 4 subunits each containing a hæme moeity. The binding and unbinding of oxygen to hæmoglobin is a cooperative process which gives rise to the following typical sigmoidal O_2-hæmoglobin binding curve, reproduced here from Figure 2.6 on page 64:



The total quantity of oxygen delivered to a tissue per unit time is equal to:

$$\begin{aligned} O_2 \text{ Delivery (ml/min)} = {} & [\text{Hb}](\text{g/ml}) \times (\text{Hb } O_2\text{-binding Capacity (ml/g)}) \times (\%\ O_2 \text{ Saturation}) \\ & \times (0.0031 \cdot \text{Partial Pressure of } O_2 \text{ (mmHg)}) \times (\text{Flow Rate (ml/min)}) \end{aligned}$$

Design Calculations

We shall discount the ≈0.3% of the oxygen supply carried in solution at normal atmospheric pressure and use a total hæmoglobin content of 70g/L from the perfusate design discussions in Section 4.6.1 on page 122. The O_2 capacity of hæmoglobin is given as 1.34mlO_2/g Hb, a value termed *Hüfner's constant*. Caveats will be discussed later on page 278 in Chapter 8, but this value will suffice for determining the requirements of the oxygen sensing system in the remaining part of this discussion.

The perfusate provides a calculated total carrying capacity of ≈93ml O_2/L at 99.9% saturation, delivering some 46.5ml O_2/min to the kidney at a flow rate of 500ml/min, from which it is known to extract a maximum of ≈11.2ml O_2/min to meet its metabolic requirements[19, 56].

Established clinical devices are capable of detecting the oxygen saturation of blood to within 1% between 75% and 99.9%[1]. Matching this degree of accuracy for the O_2 saturation sensor would allow us to determine the O_2 content of the perfusate to within:

[1]The reason for this particular range of calibrated saturations in clinical devices derives from the methodology of in-vivo calibration and the danger of reducing blood oxygen saturation levels below 75% in study patients.

$$\begin{aligned} x \text{ ml}O_2 &= \text{Hæmoglobin} \times \text{ Hüfner's constant} \times 1\% \\ &= 0.938 \text{ ml}O_2 \end{aligned}$$

Or 8.4% of the maximum 11.2ml O_2 consumed by a kidney each minute. This degree of accuracy is reduced if the kidney actually consumes less O_2, or the true Hüfner's constant is higher, but improves if the total hæmoglobin content is reduced.

As an indication of the results to be expected from the system, 11.2ml of O_2 consumed from the perfusate would reduce the O_2 saturation from 99.9% in the arterial limb to 76% in the venous limb. With the hæmoglobin content under the control of the system designer, detecting the absolute hæmoglobin content is only a secondary goal of the sensor system at this stage.

The bandwidth requirements of the oxygen sensing system are dictated by the rate of delivery and consumption. Under normal conditions, the rate of change of renal oxygen consumption should be near zero as its work remains constant. We could expect a failing kidney to slowly change its rate of consumption over a number of minutes to hours, requiring a bandwidth of only 0.1Hz to capture this decline.

I can also imagine scenarios where the detection of sudden hypoxic or anoxic insults such as those resulting from embolic events would be of great benefit, and a target bandwidth of 1–2Hz should satisfy this use case.

Final Design Targets

Range: 75% – 99.9% O_2 saturation Sensitivity: 1% Bandwidth: 2 Hz

5.5.4 Flow Rate

The primary goal of the flow sensor is to work in combination with the oxygenation sensor to derive the momentary oxygen consumption rate of the kidney. If we wish to monitor tissue O_2 delivery to match the accuracy of the O_2 sensor as calculated on page 164, we require an ability to distinguish step changes of ≈1ml O_2/min. From this we can calculate the required sensitivity of the flow sensor:

$$1.34 \text{ mlO}_2\text{/g Hb} \times 70 \text{ g/L} \equiv 0.0938 \text{ mlO}_2\text{/ml perfusate}$$
$$\therefore 1 \text{ mlO}_2\text{/min} \equiv 10.66 \text{ ml/min}$$

Selecting a sensitivity of 10ml/min would thus be a reasonable design target, and applied to the normal range of biological flow rates discussed previously under the design of the Tesla pump (page 122 onwards), with a degree of headroom.

The bandwidth of the flow sensor should match or exceed that of oxygenation sensor with which it will be paired, possibly extending up to that of the pressure sensor to capture the entire cardiac cycle. However, in the initial embodiment of the device, a lower bandwidth of 1–2Hz would still match the functional requirements.

Final Design Targets

Range: 0 – 600 ml/min Sensitivity: 10 ml/min Bandwidth: 2 Hz

Benefits of Differential Measurement

Differential measurements between the arterial and venous limbs of the circuit allow for the derivation of useful physiological data:

- Differential temperatures can reflect thermogenic metabolic activity
- Pressure drops can reflect resistance to flow in the vascular bed and therefore sludging or microvascular stasis
- Differential flow rates can be used to identify leaks within the system, and possibly the rate of urine production
- Differential oxygen saturation measurements simplify the derivation of oxygen consumption by the tissue

5.6 Summary of Design Requirements

The following table contains all the requirements for the SmartPipe's sensor suite established in this Chapter:

	Temperature	Pressure	Flow	Oxygenation
Range	0 – 40°C	0 – 200 mmHg	0 – 600 ml/min	75% – 99.9%
Sensitivity	0.1°C	2 mmHg	10 ml/min	1%
Bandwidth	1 Hz	200 Hz	2 Hz	2 Hz

Table 5.1: Summary of the design requirements for the SmartPipe's sensor system

Additionally, the final design of the pipe had to have an internal diameter of 6mm in order to match the normal renal artery.

5.7 Design and Development Process

Achieving the multiple requirements for the SmartPipe system in the short time available required a rapidly iterative design and prototyping methodology for the circuit schematics, PCB layout, programming and 3D design of the through-flow pipe to interface with the sensors.

This process took the following form:

1. Review literature
2. Sketch circuit concept
3. Source components from commercial supplier
4. Simple PCB layout to determine component placement
5. Mill out single-layer test board
6. Surface-mount components
7. Write code to interface with component
8. Test component
9. Iterate design

And once a final design had been solidified:

1. Convert PCB layout to smaller multi-layer design
2. Send PCB for professional manufacturing
3. Test PCB with previous code
4. Iterate code and design

Some of these stages will be covered in the rest of this section.

Selection of the Embedded Microcontroller

One of the first decisions I made in the development stage was to settle upon a reliable microcontroller platform which would allow me to interface with and assess the many

circuit designs and integrated chips selected for potential use in the system over the duration of the project.

The distinction between microprocessor and microcontroller has grown over the decades, with microprocessors now being in the minority and dedicated solely to 'number crunching' applications, with I/O functionality limited to communication with fast memory or essential peripheral controllers.

A microcontroller, on the other hand, is a fully integrated device containing hardware for data storage and manipulation, I/O for interfacing with external devices using digital communication protocols, internal power management and other functions specific to the device and its family, often including ADC or DAC functionality, user interface control, visual display capabilities, and occasionally integrated wireless communications.

Design Requirements

Developing the SmartPipe concept required an integrated microcontroller able to perform most of the following:

1. Manipulate digital data with simple mathematical routines
2. Interface with external user inputs e.g. buttons
3. Communicate with other chips using standard digital communication protocols:
   (a) Analog interface chips (ADCs and/or AFEs)
   (b) External DACs for transmitting processed data
   (c) Wireless controllers (in future versions)

In addition to these functional requirements, the physical requirements included:

1. Availability
2. Low cost
3. Small surface area required
4. Low power consumption suitable for a battery-powered device
5. Possible multiple power-modes and sleep functionality for inactive periods
6. Simple power supply requirements
7. Free software development toolkit
8. Online resources including discussion fora and datasheets

Selection of the ATMega328P

I selected the ATMega328P[1] because of a combination of personal experience with the platform and its ability to satisfy nearly all of the stated design requirements for the SmartPipe except for integrated wireless functionality.

[1] Previously from Atmel Corporation, San Jose California. Acquired in 2016 by Microchip Technology, the manufacturers of a previously competing line of 'PIC' microcontrollers

The ATMega328P is an 8-bit microcontroller with 32KB of on-chip flash memory for program storage, 2KB of RAM and a 1KB EEPROM for holding long-term data between power cycles such as calibration coefficients and look-up tables. The controller has access to 32 external pins, with 24 of these connected to interrupt functions including wake-from-sleep. There are also 3 internal timers (one 8-bit and two 16-bit resolution), 5 software-selectable power saving modes, and hardware communication protocols including the common *Inter-Integrated Circuit* (I^2C) and *Serial Peripheral Interface* (SPI) protocols.

The controller requires only 3 pins to be connected to a digital supply of between 1.8V – 5.5V for operation without the on-board 10-bit ADC. The device can also accommodate a range of external clock supplies, including crystal, oscillators and resonators to operate between 125kHz – 16MHz, with power savings at lower speeds. Clock rates over 8MHz are only achievable with $V_{SUPPLY} > 3.3V$.

Because I did not envisage any timing loops requiring the tight tolerances of a crystal oscillator, I elected to use a smaller and cheaper surface-mount resonator at 8MHz and power the system at 3.3V to reduce the current supply and to match the voltage requirements of many other modern ICs.

Furthermore, the ATMega328P is available in a wide variety of packages at a cost of $<$ £5 per unit from all major component suppliers.

Software Development Environment

The ATMega328P is better known as the microcontroller used by the Arduino project. The commercial success and widespread popularity of the Arduino platform has made software development for the ATMega328P far more simple than that for competing platforms, including the difficult Java environment for Microchip's devices (MPLAB® X), the variety of IDEs available for the MSP430 from Texas Instruments or the costly tools for ARM development. As a result of this popularity, there is a large support network available for development and debugging problems, and a large number of code libraries and samples available for a wide variety of functions and external ICs.

The Arduino Integrated Development Environment (IDE) incorporates all functionality required to develop and deploy software on ATMega328P-based hardware. This includes cross-platform support (Windows, Mac OS X and Linux), a C++ compiler, a protocol for loading compiled byte-code over a serial interface or SPI, and the ability to upload new bootloader software to the chip to modify clock rates. All of these software loading functions can take place with the chip already deployed and mounted in the destination hardware, greatly simplifying deployment.

All software development for the SmartPipe concept took place in Arduino IDE 1.0.2 and above, and was coded in C++. Interfacing with the majority of the external chips used in the project required me to write new function libraries which will be open-sourced after the submission of this Thesis. A small number of other devices could be used with minor modifications to existing open source libraries, with my work being fed back to the developers of those libraries. The specific design of the SmartPipe software will be covered in greater detail from page 206 in Chapter 7.

Further references to the ATMega328P microcontroller will refer to the microcontroller as the *ATMega328P*, and the software or the development environment as the *Arduino*.

PCB Design and Layout Software

All PCB design work was carried out in EAGLE versions 6.6 – 7.6[1] on a 2013 Apple MacBook Pro 2.7 GHz Core i7. The software was under a 'Make Personal' license, allowing multiple schematic pages, larger PCBs (up to 100×160mm) and up to 6 PCB layers. I performed all routing manually throughout the project.

EAGLE was able to deliver Gerber files suitable for the department's single-layer milling machine as well as professional manufacturing by an external company.

Breadboard Milling Machine

The department's LPKF H60 CNC machine[2] was suitable for rapidly prototyping single-layer PCB designs.

The manufacturer's data sheet claims 1μm resolution but unfortunately the real-world performance was far below this. It was however suitable for milling long pads down to $\approx$500μm separation, and tracks down to the same width. This prevented me from rapidly prototyping most components in high pin-number Quad Flat No-Lead (QFN) packages but not Small Outline Integrated Circuit (SOIC) or Quad Flat Packages (QFP) with exposed pins. See Figure 5.6 on page 171 for an example of a component mounted onto one of my custom-milled prototyping boards.

Surface Mounting of Components

Electing to use surface mount components allowed me access to a wider range and selection of chips for testing and development at much lower cost than their larger counterparts

[1] Originally CadSoft Computer GmbH, now part of AutoDesk, Inc., San Rafael, California, USA
[2] LPKF Laser & Electronics Ltd., Garbsen, Germany

owing to general commercial availability. This also reduced the iterative steps required to shrink the board layouts before sending out for professional manufacturing.

Surface mounting typically involves the use of a specially-formulated solder paste composed of flux and microscopic solder beads in suspension, allowing components to be placed onto the circuit board and handled for a short period of time without any dislodgement. This mounting is usually carried out by an automated pick-and-place machine, before feeding the completed circuits into a special *reflow oven*, capable of reaching precisely controlled temperatures at specific times. The standard *reflow profile* is shown below:

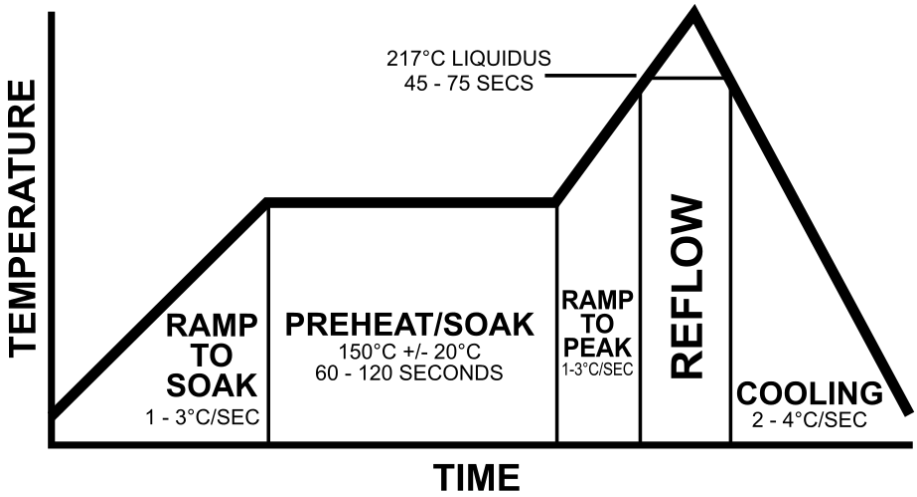


Figure 5.5: A typical time-temperature profile for reflow soldering

As one can see, there are a number of different stages in this process – initial heating of the PCB to around 150°C and holding to ensure the entire PCB rises to a uniform temperature, before ramping up to 180–225°C in order to evaporate off the flux within the solder paste, liquify the solder particles and ensure they form good connections between the components and their mounting pads.

Following this, a controlled cooling phase ensures even shrinkage of the entire circuit to prevent solder cracking or dismounting of components. The precise temperatures of the profile depend on the formulation of the paste, but the general schema remains the same for all.

I was able to source a low temperature solder paste (185°C peak) and with the aid of a binocular microscope became capable of placing surface mounted passive components such as resistors and capacitors down to 0402 format (1mm×0.5mm), and surface-mount ICs down to 40-pin QFNs and 9-ball BGAs. These were then subjected to a manually-controlled 'reflow profile' in a small recirculating fan oven. All prototypes and final PCBs were manufactured in this way.

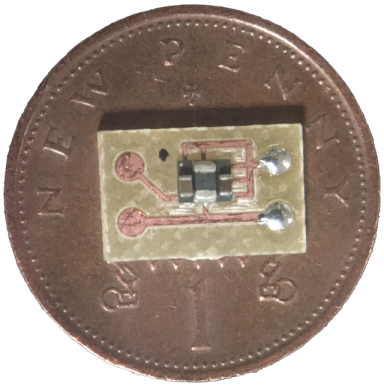


Figure 5.6: Image of a milled prototype PCB with a surface mounted TPS82740B in 9-pin BGA format, which is discussed in more detail on page 202

Professional PCB Manufacture

All professional PCB manufacture was outsourced to Elecrow Bazaar, based in Shenzhen, China. They were able to produce 4-layer boards from 0.8mm – 2mm thickness in standard FR-4 material with PCB tracks of just 200μm, and vias down to 300μm at a cost of <£10/unit, including 10-day international shipping.

This greatly assisted the prototyping stage for smaller components beyond the capabilities of the milling machine, such as the AFE4490 in its 40-pin QFN package (discussed later from page 304 in Chapter 9).

3D Design and Prototyping

As with the prototype Tesla pump (page 128), all 3D design work was undertaken in SolidWorks® Education Edition 2013-2014[1] on a 2013 Apple MacBook Pro 2.7 GHz Core i7 running Windows 7 in Boot Camp. This generated high-polygon stereolithography (STL) files which were then taken to one of the department's 3D printers depending on the degree of precision required. Both printers were able to produce objects up to around 20cm$\times$20cm$\times$15cm in size.

Low-resolution and simple designs such as that for the DAC housing (shown later on page 189 in Chapter 6) were 3D printed on Ultimaker 2 printers[2]. This 3D printer uses a spool-fed ABS plastic filament system which is melted in the delivery nozzle and laid down with a resolution of $\approx$400μm in all planes.

A second 3D printer, an Objet30™ [3], was used for more detailed designs such as the

[1]Dassault Systèmes, Véliz-Villacoublay, France
[2]Ultimaker Ltd., Preston, Lancashire
[3]Stratasys Ltd., Eden Prairie, Minnesota, USA

contoured pipe for mating with the sensors. This printer uses a layered photo-activated polymer system that simultaneously produces both the supporting structure and the hard polymer body, allowing for more complex 3D structures with a resolution of $\approx 100\mu$m in all planes.

The next Chapter will discuss the data collection system and the design of the DAC board in more detail.

Chapter 6

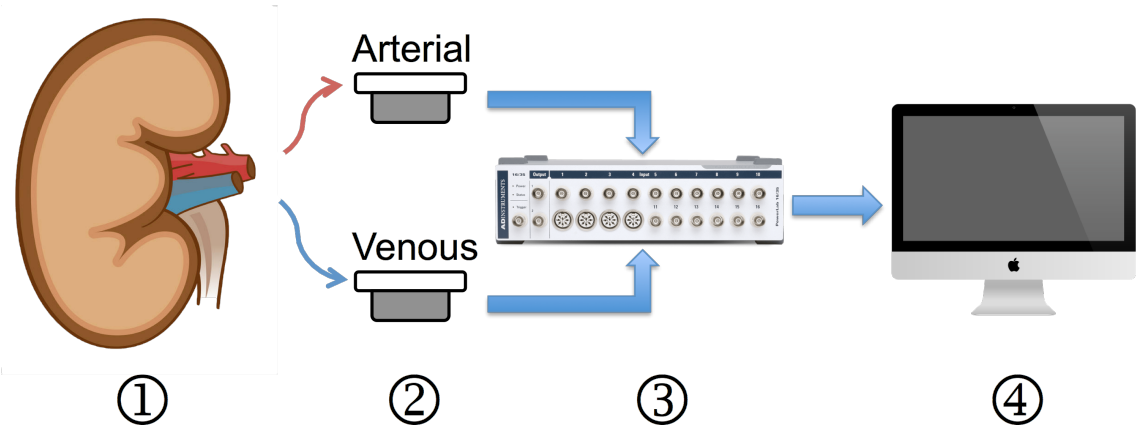
The Data Collection System

This Chapter describes the design of the data collection system for developing and testing the SmartPipe.

6.1 The Hardware Signal Chain

Prototyping the various analog sensor subsystems for the SmartPipe both in terms of electronic circuitry and C++ code required a means of relaying and capturing digitised data into a visual format more appreciable to human interpretation than a scrolling sequence of values over a serial interface.

Furthermore, the ability for future researchers to collect, store and analyse analog data arising from the various sensors of the SmartPipe was a key goal of this project, represented by the data acquisition system in stages 3 and 4 of the figure below:



6.2 Selection of Data Acquisition System

Our research laboratory has many years of experience with LabChart[1], a commercial software package which is highly optimised for collecting, manipulating and analysing multichannel time-series data. Continuing with the use of this software package was therefore the quickest way to begin development on the SmartPipe system, and all remaining data collection work in this section of the Thesis was undertaken with LabChart 7.2.5 for Mac OS X.

However, one important caveat to this decision is that LabChart will only accept data received from a licensed PowerLab data acquisition device from the same company.

[1] ADInstruments, Dunedin, New Zealand

Interposition DAC

To take advantage of the capabilities of LabChart for the rest of the project, I therefore decided to design and build a multichannel Digital-to-Analog Converter (DAC) to take digitised signals directly from the ATMega328P microcontroller and replay them into a 4-channel PowerLab/4SP system.

This system would be used for the rest of the project for both prototype boards and the final SmartPipe, requiring only the use of a portable C++ software library I wrote for the DAC.

Whilst not a problem during the prototyping stage, one significant concern was that this additional step would contaminate the otherwise 'clean' signal chain in the final SmartPipe device, shown in Figure 6.1 below:

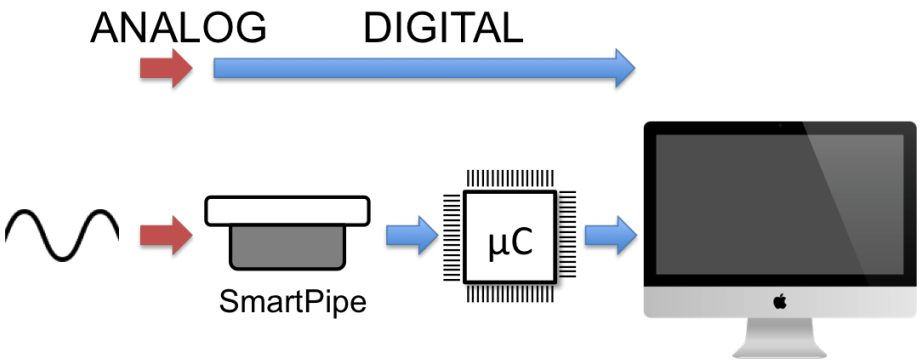


Figure 6.1: Image of a 'clean' analog to digital signal chain with a single-stage conversion. The red arrow indicates susceptibility to noise

In this figure, the front-end electronics of the SmartPipe have been highly optimised to collect and digitise data from the analog sensors with as little noise and interference as possible. This is then relayed all the way through to the data visualisation software as high-fidelity digital data which is immune to noise.

Signal Contamination

Using an interposition DAC to interface with a PowerLab device 'contaminates' the digital data collected from the SmartPipe. This comes from the DAC →ADC conversion step as well as the lower specifications of the capabilities of the DAC and PowerLab versus those of the SmartPipe's analog front end. This contamination is shown in Figure 6.2 below:

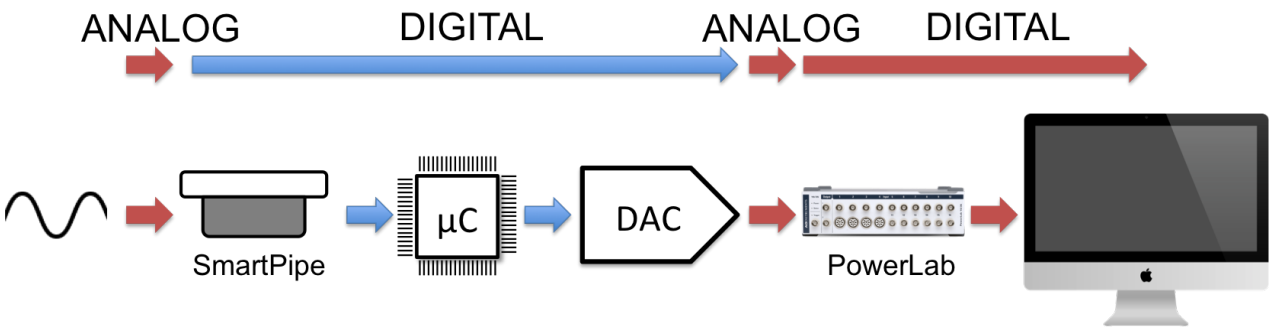


Figure 6.2: Image of the 'contaminated' signal chain. The red arrows indicate noisy or error-prone signals

As the sole means of data collection and visualisation, it was not possible to know the degree of signal degradation that this extra DAC→ADC stage caused to the original signal beyond the published specifications of each device, discussed from page 183.

6.3 Minimising Conversion Errors and Noise

There are certain concepts relating to signal fidelity and electronic noise which came into my decisions during the design of the DAC and other stages of the SmartPipe. They will be outlined here in brief.

Conversion Errors

Conversions between digital and analog signals are dependant on the quality of the *transfer function* of the relevant ADC or DAC device. One such transfer function is presented in Figure 6.3 below:

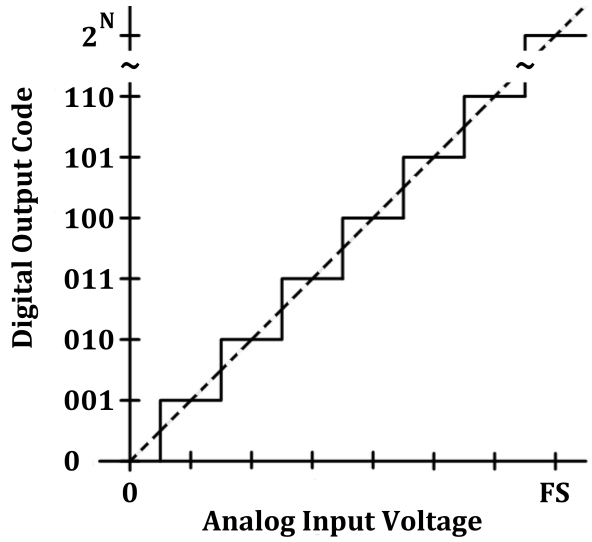


Figure 6.3: Transfer function of an ideal ADC as the dashed line, with a real-world perfect ADC demonstrated by equal-sized steps for each change in analog input voltage and a corresponding digital output code. FS = 'Full-Scale' Input

In this figure, the dotted line represents the ideal $x = y$ slope wherein the device is able to resolve the fully continuous analog signal into an infinitely small number of digital codes.

The superimposed step function illustrates the real-world correspondence between the analog input voltage (x axis) and quantised digital output codes (y axis). Note how the horizontal span of each digital code extends to ± 0.5 the minimum voltage step size around the target voltage.

The number of steps along the vertical span corresponds to the resolution of the ADC – how many binary steps it is able to resolve, or how finely *quantised* the input can be. Resolutions commonly range from 8 – 24 bits (–1 LSB) for typical ADCs, indicating $2^8 - 1$ = 255 to $2^{24} - 1$ = 16,777,215 steps along the full span.

Given a perfect 4.096 reference voltage to provide the full-scale point along the x axis, an 8-bit ADC would be able to resolve analog steps down to $\frac{4.096\text{V}}{2^8 - 1} \approx$ 16mV, a 16-bit ADC 62.5μV, and a 24-bit ADC just 244 nanoVolts. However, no sufficiently accurate standalone references exist for >16-bit ADCs, requiring ratiometric referencing which will be covered later on page 233 in Chapter 8.

Non-Linearity – INL and DNL

The quality of the signal converter's transfer function is reflected in its *Differential Non-Linearity* (DNL) and its *Integral Non-Linearity* (INL).

The DNL describes any deviation from the ideal width of a conversion step. This error arises from the actual architecture of the signal converter in silicon, and is therefore shared

amongst families of similar devices and cannot be compensated through calibration.

The DNL errors integrate along the entire input range of the device and lead to deviations in curvature of the slope of the transfer function, giving the INL. This is measured at the point of *maximum* deviation of the transfer function from the ideal $x = y$ relationship.

High fidelity devices will typically have DNL values <0.5 LSB and INL values of <2 LSB.

Noise

Effective Number of Bits (ENOB)

Unlike the INL and DNL which represent the linearity of the transformation, the ENOB is a reflection of a signal converter's noise performance. This is a way of representing the device as though it was a perfect converter of lower resolution.

For example, a perfect 16-bit ADC with a perfect 4.096V reference should theoretically be able to quantise a signal of 62.5μV into one LSB.

However, if the RMS noise on the input is 125μV, equivalent to 2 LSB, the performance of the 16-bit ADC will in fact be degraded to that of a 13-bit ADC as follows:

The IEEE-1057 Standard defines the ENOB as:

$$\text{ENOB} = \log_2 \left(\frac{\text{Full-Scale Voltage Range}}{\text{RMS Noise} \cdot \sqrt{12}} \right) \tag{6.1}$$

Thus, for the aforementioned 16-bit ADC with a background noise of 125μV:

$$\begin{aligned} \text{ENOB} &= \log_2 \left(\frac{4.096\text{V}}{0.000125 \cdot \sqrt{12}} \right) \\ &= 13.2 \text{ bits} \end{aligned}$$

Thus ENOB can be a more useful measure of the true capabilities of a signal converter for a given noise environment.

6.4 Principles of Component Layout

After designing a functioning circuit and selecting appropriate components, the next stage is to settle upon an appropriate physical PCB layout. This is based on design experience and an understanding of the principles behind signal noise in order to maximise performance.

The major mechanism shared by all noise sources is energy transfer between components and traces through the electromagnetic field – principles first explored by Michael Faraday (1791 – 1867) in the 1830s.

A circuit designer intends for all currents to remain confined to specific traces on the board and travel only in a point-to-point fashion where guided. However, electrons in motion induce surrounding magnetic fields, and these magnetic flux lines are then able to transfer energy and induce the motion of electrons in any other conductors through which they pass.

Capacitive Coupling

In its basic form, a capacitor is a pair of conductive plates separated by a dielectric, including any two conductive traces on a PCB separated by the fibreglass substrate.

The resulting unwanted, parasitic capacitance is a combination of the area of conductor overlap (A), the separation between the conductors (d), and the dielectric value (ε) of the interposed material:

$$C = \frac{\varepsilon \cdot A}{d}$$

Parasitic capacitances can even develop between the pads of an IC and board, resulting in values of just a few femtoFarads to picoFarads, yet still enough to interfere with the transition times of high speed signals – usually considered to mean >10MHz, which is sufficient to include clock lines and high speed digital communications.

Inductive Coupling

This effect arises from any current-carrying trace. As electrons travel along the trace, a magnetic field emanates from it proportional to the magnitude of the current. The electrons of any conductor sufficiently close to fall within range of the emitted EM field will receive energy through this field and begin to move, resulting in an induced current.

Because flux lines are only ‘cut’ by time-varying currents, high speed and switching traces are again more susceptible to this effect than DC ones.

Given the magnetic field strength (B) between two parallel wires in a vacuum separated by distance r, carrying current I and the magnetic permeability μ_0:

$$B = \frac{\mu_0 \cdot I}{2\pi r}$$

Resulting in a magnetic flux (Φ):

$$\phi = B \times \text{Area}$$

By Faraday's law:

$$\text{Induced e.m.f (V)} = -\frac{\partial \phi}{\partial t}$$

Thus, decreasing the signal rate of noisy digital traces, or reducing the area susceptible to a stray EM field by increasing the physical separation of traces and reducing the trace length are all ways of preventing this type of coupling.

The Role of Impedance

Finally, the concept of impedance explains why certain components and traces are more susceptible to noise than others.

As per Ohm's law, the voltage induced by a current will be proportional to the impedance of the circuit ($V = I|Z|$) and vice versa. Thus small induced or coupled currents can result in large noise voltages into circuits with very high impedances, such as unbuffered analog signal lines.

Design Principles

Given the methods by which unwanted noise couples into components and traces on a board, it is important to be mindful of the following design principles:

Reduce Impedances

Lower impedances are less susceptible to noise. This means:

- Using short, fat traces instead of long thin ones where possible
- Buffering sensitive signals as close to the source as possible
- Using ground and power planes instead of ground and power lines

Minimise Trace Lengths

This particularly applies to high impedance analog traces – the greater the length of conductor, the more likely it will be exposed to EM fields.

Use Shielded Twisted Pairs

External cables should be shielded twisted pairs where possible. Grounding or driving the shield can reduce the noise even further.

Minimise Crossover

Traces should always cross at right-angles to each other so that the shared area of overlap is minimised, thereby minimising capacitive coupling. This particularly applies to running potentially noisy traces (digital or power lines) beneath the analog pins of any IC.

Separate Analog and Digital Signals

Digital signals are low impedance, and rapidly switch between voltages defining 0 and 1. This can often mean 8MHz – 30MHz swings between 0 and 3.3V for digital communication lines.

Physically separating the analog inputs to one part of the board, and digital lines to another is one key way of avoiding any inductive crosstalk. Susceptible analog regions should be defined and kept 'sterile'. Even digital signals routed under analog regions or components can lead to induced return currents interfering with the analog signals on the other side of the board.

Using separate voltage regulators to provide the analog and digital power supplies also reduces a possible avenue of 'contamination' between these signal domains.

Avoid Loops

Loops act as antennas and inductors. Any time a return current has to take a non-linear path to the point of lowest impedance (ground), a current loop can develop and emit or collect EM energy.

These can take the form of long physical traces routed around the edge of the board in order to avoid crossing other components, or ground planes cut by either physical slots or other traces on the plane.

Use a Large Ground Plane

Low impedances will 'attract' the most stray EM energy. For this reason, providing a large ground plane maximises the absorption of stray EM fields and allows for straight-line ground return paths beneath conductive traces. Multiple ground points for sensitive components, as well as interposed ground lines ('guard traces') between regions susceptible to crosstalk are another good way to reduce interference by EM field coupling.

A large ground plane also ensures that all components share a common reference point and minimises ground currents. It can also act as a heat sink for any high power components, keeping in mind that uneven heat distribution between different materials (e.g. copper ground planes and tinned solder pads) can create minor voltage perturbations of their own due to the Seebeck (thermoelectric) effect.

Decouple all Power Pins

Providing capacitors close to IC power pins provides a low impedance route for power line noise to return to the ground plane rather than entering the IC, as well as acting like small, low-impedance 'batteries' to compensate for any voltage drooping which might otherwise interfere with their performance.

Capacitors should ideally have low *Equivalent Series Resistance* (ESR) and very high leakage resistance – ceramic capacitors satisfy both conditions with ESRs of milliOhms and leakage resistances $\geq 10^9 \Omega$.

Use Multi-Layer Boards

Multi-layer boards allow for the best control of physical separation of different signal types so as to prevent cross-talk. They also provide a way to place low impedance ground planes directly beneath important pads and traces to reduce coupling and noise, and can separate sensitive components from a potentially noisy power plane.

One recommended 'stack-up' of a 4-layer board is illustrated below:

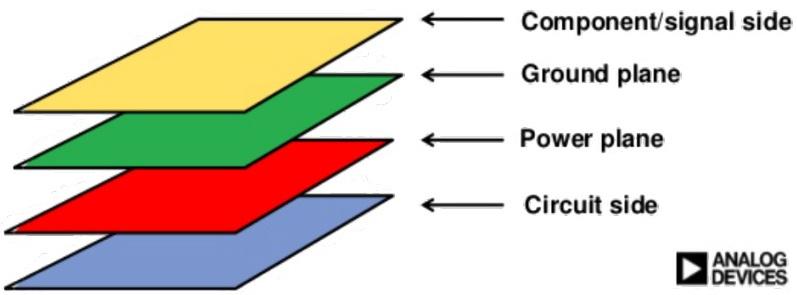


Figure 6.4: Illustration of a 4-layer PCB. Analog Devices Inc.

This layout places all sensitive analog components on one side of the board, and noisy digital traces on the other. Interposed between these two are a ground plane immediately beneath the sensitive top layer, and a power plane closer to the noisier digital bottom layer.

6.5 Building the DAC Unit

With these design principles in mind, the DAC was implemented as a modular multi-layer PCB design, housed within a custom-designed and 3D printed enclosure. It was powered from a 6V wall-mounted DC adapter and capable of generating 4 simultaneous output signals to feed into a multichannel PowerLab/4SP device.

Selecting the DAC

Accommodating the desired signal outputs of temperature, pressure, oxygen saturation and flow rate required the selection of a multi-channel DAC that balanced a combination of performance, signal range and cost, as well as the capabilities of the PowerLab/4SP.

The PowerLab/4SP provides 4 inputs (single-ended or differential) with multiple range settings and a maximum 40kHz sampling rate for 4 simultaneous inputs. The manual claims 16-bit resolution with ± 1 LSB INL and $<1\mu V_{RMS}$ noise.

The DAC selected was the DAC8564 from Texas Instruments[1], a 16-bit, 4-channel device capable of achieving a 10MSPS rate with an output range of 0–V_{REF}, ≤ 4 LSB INL and 0.25 LSB channel-to-channel crosstalk. This was available in a 16-pin surface-mount package at a price of around £12/unit.

In the worst-case scenario, the capabilities of this DAC would be the limiting factor in signal fidelity over the PowerLab.

[1]Dallas, Texas, USA

This is unfortunate because the design of the SmartPipe front-end would have captured the signals from the analog sensors with $\leq$2 LSB accuracy (covered later in Chapter 9), but this was the best quad-output DAC available in its price range.

Thus there was a significant risk that initial 15/16-bit accurate data would be degraded to a signal with a lower ENOB once in LabChart, but there was no real way to work around this compromise during the prototyping stage.

Overcoming the Limitations of the DAC

One way to work around this signal degradation in the finished device was to capture and then scale or convert the data into a representation of its final magnitude within the SmartPipe itself, spreading the high fidelity data across a greater range of LSBs in the final digitisation by the PowerLab.

For example, a temperature increase from 20.0°C to 20.1°C is a step change of just 0.0389Ω in the total resistance of the RTD (page 232), or 19μV with an excitation current of 500μA (see page 226 for a discussion of RTD measurement methodology). However, representing this 20.1°C signal as an output of 2010mV would provide 10mV separation between fractional steps from 0°C to 45°C, requiring only ≈9-bit accuracy with a 5V reference.

Similarly, a post-calibration pressure of 120mmHg could be represented as a signal of 2400mV, thus providing 40mV per 2mmHg-significant step from 0mmHg to 200mmHg at 4V. This translates the sensitive output of the front-end pressure sensor to signal requiring only ≈7-bits with a 5V reference.

Reference Selection

Running on its 2ppm precision internal reference, the DAC8564 is only capable of delivering a signal range of 0–2.5V. I decided to provide the DAC with an external reference to increase the signal range to 0–4.096V. In retrospect, this change was unnecessary as the PowerLab would have simply been able to accommodate the 0–2.5V range by selecting the appropriate internal range in software.

The reference used was the REF5040 from Texas Instruments, a high precision series resistance device specified for 16-bit performance in a miniature MSOP-8 surface-mount package.

- Temperature drift: 3ppm/°C
- Initial Accuracy: ±0.05%
- Noise: 12μV_{pp} from 0.1Hz – 10Hz
- Stability: 5ppm/1000hr

Power Regulation

The DAC board was designed to be powered from a small wall-mount AC/DC adaptor providing 6V – 9V at 500mA. The power entered the board through a polymer positive temperature coefficient (PTC) fuse rated at 500mA. This resettable fuse was designed to protect against surges and a malfunctioning AC/DC adaptor.

On the board itself the power was split into +5V and +3.3V rails via MIC5205 Low Drop-Out linear regulators (LDOs) in order to power the DAC8564 and an Si8640[1] digital isolator, respectively. The power lines were run with large surface traces rather than planes because of the low component count and simplicity of the point-to-point layout.

Electrical Isolation

Ferrite beads were placed on all input and output communication lines from the board to resist EMI, particularly when external connections are made with the board.

Because the final working environment of the SmartPipe were not fully known, I adopted an over-cautious approach to the electrical isolation of the design. I targeted the voltage isolation requirements of the IEC60601 (3rd edition) standards for medical electrical equipment, placing an Si8640 high-voltage digital isolator on the DAC's SPI communication lines.

The Si8640 uses a unique on-chip RF induction technology to provide 4 unidirectional digital communication channels at rates up to 150Mbps whilst maintaining an isolation barrier rated at $5kV_{RMS}$.

However, in order to maintain isolation, both sides of the barrier require their own power supplies, and each channel draws between 1.6mA–5.5mA during operation. This was an excessive current draw with the SmartPipe designed to operate from a small battery, and so the Si8640 was not used in the final laboratory-based design, although it was laid out on the PCB.

The ground planes of the multi-layer board were also separated according to the isolation domain in order to prevent potential cross-talk or breakdown of the $5000V_{RMS}$ isolation barrier. These separate ground planes are shown in Figure 6.5 below:

[1] Silicon Labs Inc., Austin, Texas, USA

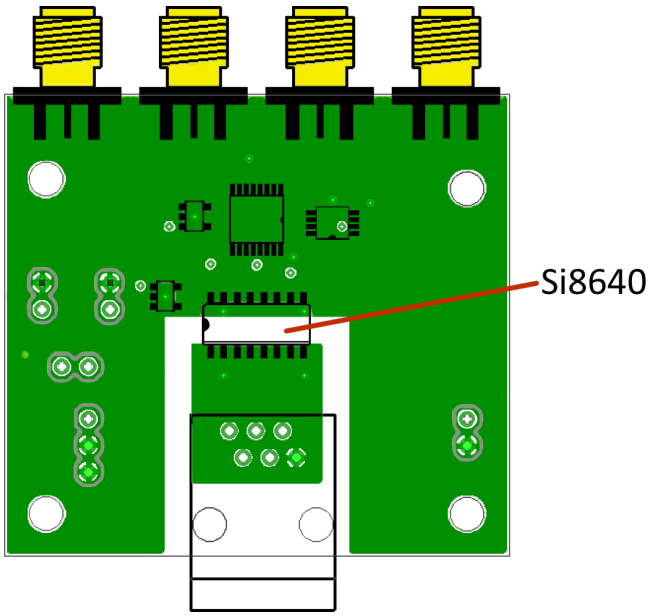


Figure 6.5: Two separate ground planes within the DAC board to ensure 5kV_{RMS} isolation across the Si8640 which isolates communication between the SmartPipe and the DAC8564

Communication with the Microcontroller

The DAC8564 is an SPI slave device, although with only 3 of the 4 lines in use (`SCLK`, `MOSI` and $\overline{\texttt{CS}}$) because it does not need to return any data back to the master microcontroller. `LDAC` and $\overline{\texttt{ENABLE}}$ were both tied to ground to force the DAC to immediately update its outputs upon receipt of new data.

The final design of the DAC used a 1-metre long wired connection to link the DAC and the ATMega328P master controller from either the prototyping boards or the final SmartPipe. This was terminated in a simple 6-wire RJ-14 telephone-style connector.

Designing for Modularity

Each PCB was designed to accommodate four outputs from one SmartPipe, but in order to allow multiple SmartPipes to operate in parallel at different key points around a perfusion circuit, the PCB design was made as modular as possible so that a number of DAC boards could be housed together and share a single power supply, yet still be turned on or off individually.

6.6 Final DAC System

An annotated image of the final PCB is given in Figure 6.6 below.

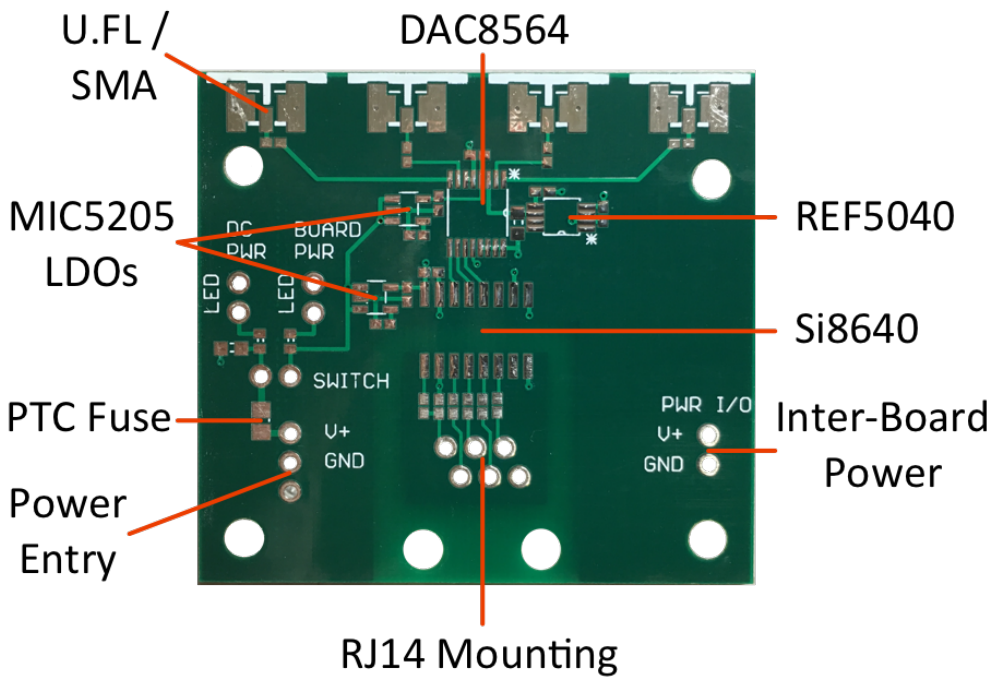


Figure 6.6: Annotated image of the finalised DAC PCB design demonstrating the key components and off-board connectors. The Si8640's pads were intentionally shorted during assembly

Two of these were placed together inside a custom-3D printed housing to allow for both the venous and arterial SmartPipes to operate simultaneously and from a single power supply.

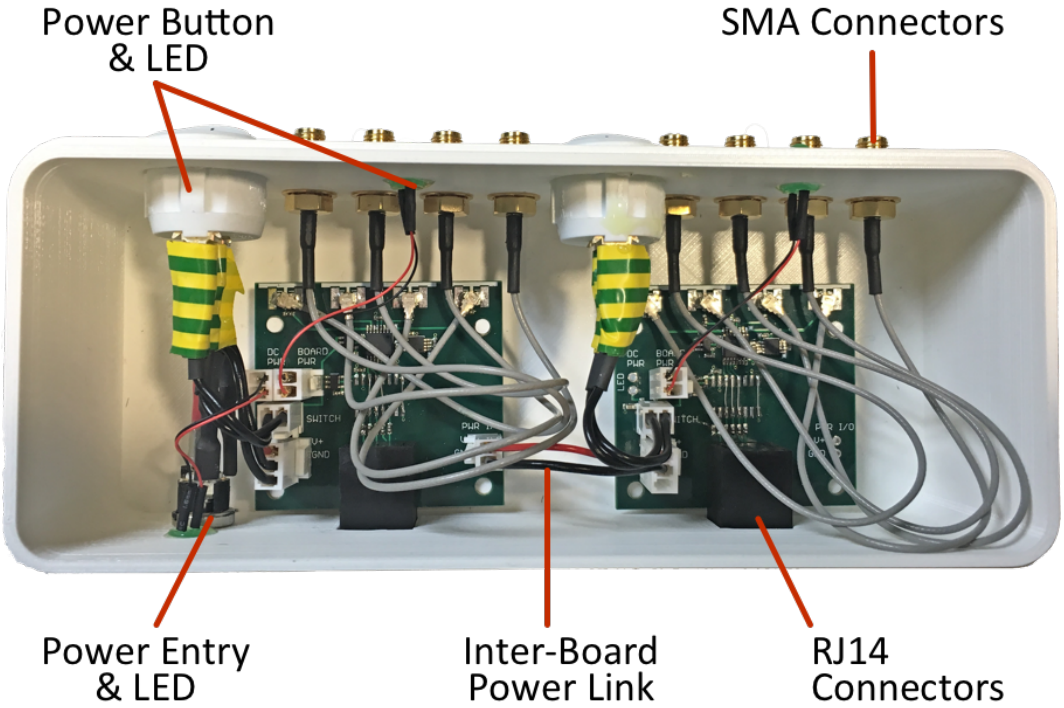


Figure 6.7: Annotated internal layout of the finalised DAC system assembled inside the custom 3D printed housing

Figure 6.7 above shows the completed board with its various off-board connections assembled inside the custom 3D printed housing, including the:

- U.FL outputs connected to standard panel-mounted SMA connections for interfacing with the PowerLab's BNC connectors
- The RJ14 connector jacks
- The panel-mount power switches and indicator LEDs
- The modular inter-board power system for powering multiple boards in parallel
- The panel-mount DC power entry jack and indicator LED

The use of the 3D-printed housing improved robustness, portability and prevented access to the powered internal components. The final system is shown in a typical testing arrangement with the PowerLab in Figure 6.8 below:

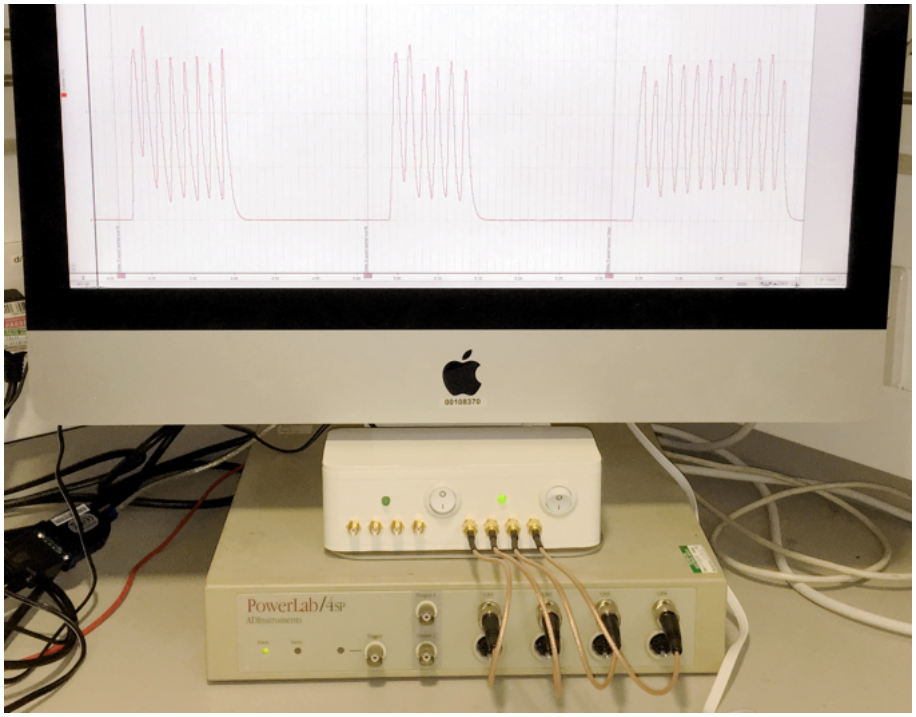


Figure 6.8: The 3D printed DAC housing beneath an Apple iMac, demonstrating the connections to the PowerLab/4SP and LabChart. The system is displaying a frequency response test of the pressure sensor

6.7 Testing the DAC

After basic testing proved that the DAC was able to convert digital codes sent by the Arduino device into analog signals for capture in LabChart, I undertook testing to derive the ENOB of the system.

This involved sending the DAC an incrementing 16-bit step function equivalent to a particular digital code resolution (i.e. steps of 0x100 for 8-bits, 0x80 for 9-bits, and so on). These results are compiled below in Figure 6.9:

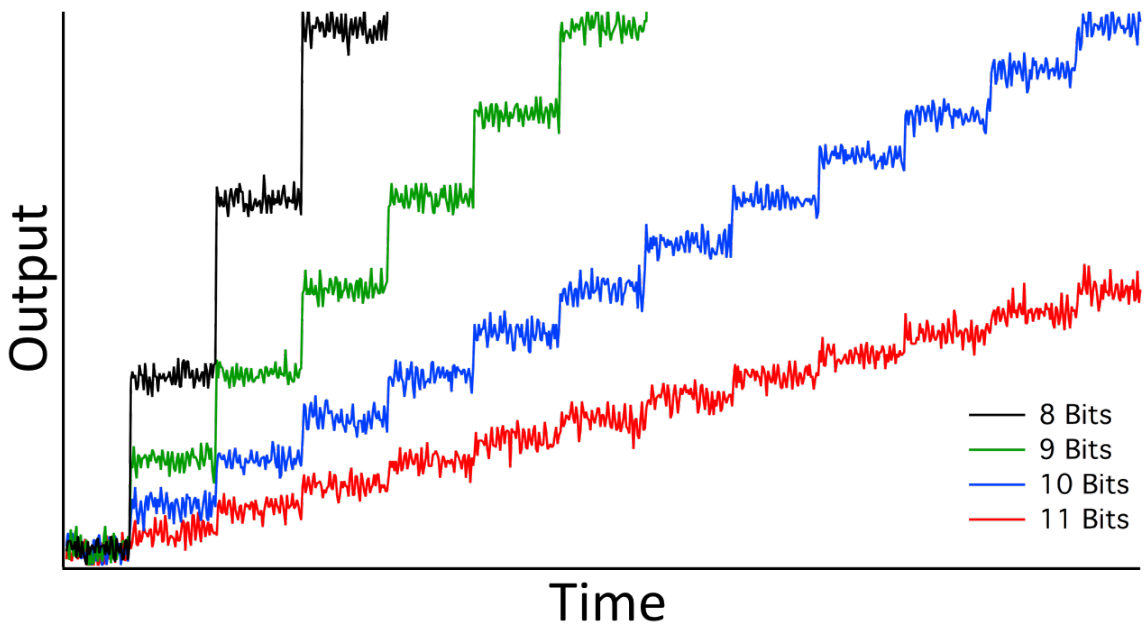


Figure 6.9: Testing the resolution of the DAC8564 + PowerLab/4SP combination. Beyond 11 bits of resolution, the steps become indistinguishable from the background noise

This demonstrated that the system is incapable of distinguishing resolutions greater than ≈11 bits. The input noise level was measured at 885μV_{RMS} within LabChart. Calculating the ENOB according to equation 6.1 on page 178 provides the following result:

$$\begin{aligned}
\text{ENOB} &= \log_2\left(\frac{\text{Full-Scale Voltage Range}}{\text{RMS Noise}\cdot\sqrt{12}}\right) \\
&= \log_2\left(\frac{4.096\text{V}}{0.000885\cdot\sqrt{12}}\right) \\
&\approx 10.4 \text{ bits}
\end{aligned}$$

Thus the output system as it stands can only be expected to perform as a system equivalent to one with ≤11 bits of resolution, confirming the chain contamination demonstrated in Figure 6.2 (page 176). Additional shielding within the housing, or perhaps driven shielding on the output lines could have perhaps improved this to some degree.

I considered this level of resolution to be acceptable for debugging and development, but a final version of the SmartPipe system would ideally stream digital data directly into a computer without passing through the interim analog conversion step. Until then, the research version of the SmartPipe would need to perform internal scaling and conversion to overcome the limitations of this particular data collection system, as described in Section 6.5 on page 184.

Chapter 7

The Bottom (Digital) Layer of the SmartPipe

This Chapter details the design and layout of the bottom layer of the SmartPipe which includes the power and user interface subsystems, as well as the custom software authored to operate the entire device.

7.1 A Master Digital Board

The introduction to the SmartPipe on page 159 illustrated the basic internal organisation of the device. This Chapter will detail the functionality of the lowermost board and the design decisions that went into its construction.

Implementing the complex multimodal mixed-signal SmartPipe required the ATMega328P microcontroller to manage a variety of digital functions such as the user interface, the power subsystem, and communications between the analog electronics and the output DAC.

I therefore decided to confine all of the digital functions of the SmartPipe to a single board that could also accommodate the potentially noisy power subsystem. This would minimise any potential contamination or cross-talk with the other, more sensitive, analog systems and allow them to be placed on their own dedicated boards in turn.

A detailed diagram outlining the functional parts of the lowermost board are given in Figure 7.1 below:

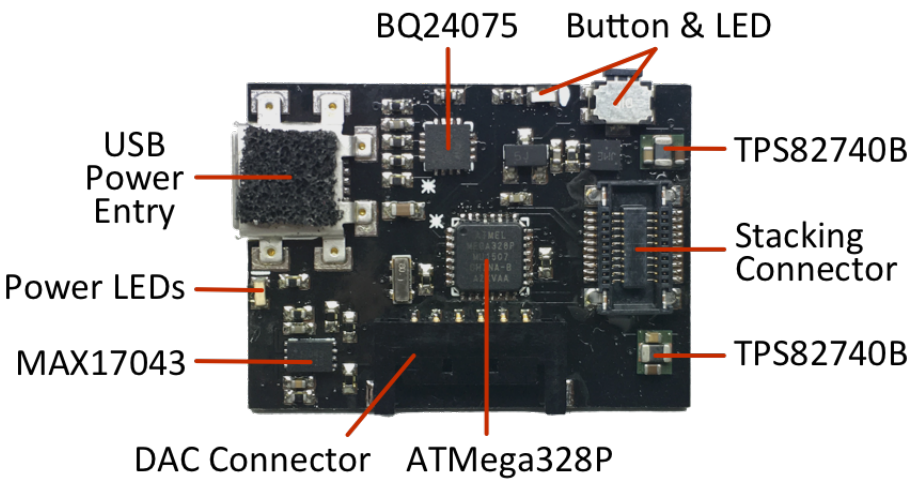


Figure 7.1: Annotated diagram of the bottom (digital) layer of the SmartPipe

7.2 The PCB and External Connectors

Applying the principles of good PCB design introduced on page 180, I designed the final digital board with the following 4 internal layers:

Top Layer:	All surface mount components
Ground Layer:	Implemented as a single plane
Power Layer:	Internally divided into multiple power supply subsections
Bottom Layer:	Solder pads for attaching the battery

All ground and power connections between the components on the uppermost layer and the relevant planes were made using 300μm through-hole vias placed as near to the target pins as possible. All power pins were decoupled at the target device with 100μF capacitors in 0402 format.

The internal layout of the power plane is better explained in the context of the power subsystem, and is shown in Figure 7.3 on page 195.

Board-to-Board Connector

A 20-pin, 500μm pitch header[1] from Molex[2] was used to relay digital communications, ground and the digital and power lines between this purely-digital PCB and the mixed-signal middle PCB. This mated with a 20-pin connector[3] mounted on the underside of the middle mixed-signal board to give a final stacking height of 2mm.

Beneath the connector, 3 separate 1mm holes were drilled through the PCB to allow the connector header to be epoxy-bonded to the PCB, physically reinforcing the system for multiple connection/disconnection cycles. Two drill-holes were used to reinforce the header on the underside of the middle board.

The detail of these connections are shown in Figure 7.2 below:

[1] Part Number: 54363-0289
[2] Molex Inc., Lisle, Illinois, USA
[3] Part Number: 55650-0288

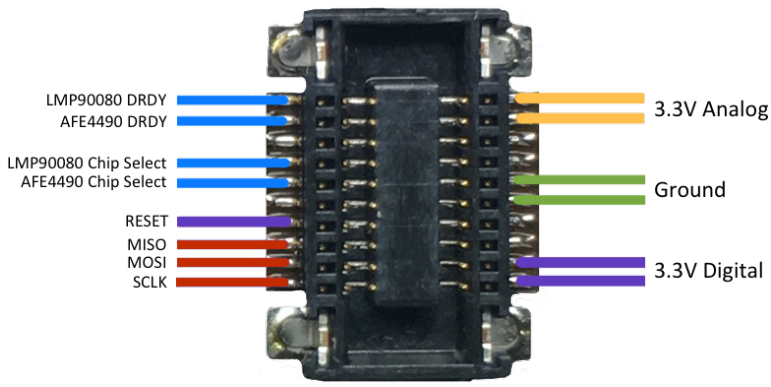


Figure 7.2: Signals carried by the board-to-board connector between the bottom and middle PCBs. Red lines carry common SPI communications, blue lines carry peripheral-specific SPI connections, purple lines carry digital power and yellow lines carry analog power

This detachable stacking method modularised the design, allowing me to iterate through various versions of each board whilst retaining common communication and power lines.

Board-to-DAC Connector

The requirements for the board-to-DAC connector were the ability to allow multiple mate/unmate cycles, with good retention of the connector once mated, and a side-entry header less than $\leq$2mm in height that was small enough to fit on the board without occupying too much surface area.

A 6-pin, 1.5mm pitch Pico-Lock™ wire-to-board header[1] from Molex was selected because it matched the stacking height of the board-to-board connection system and provided a locking press-fit mating system for the RJ14-terminated cable into the DAC board, unlike other systems which could detach with minimal force or were single-mating-only connections. This would carry a ground connection, as well as the $\overline{\texttt{CS}}$, `MISO`, `MOSI`, and `SCLK` lines.

7.3 Power Subsystem

Figure 7.3 below illustrates the design of the power subsystem:

[1]Part Number: 504050-0691

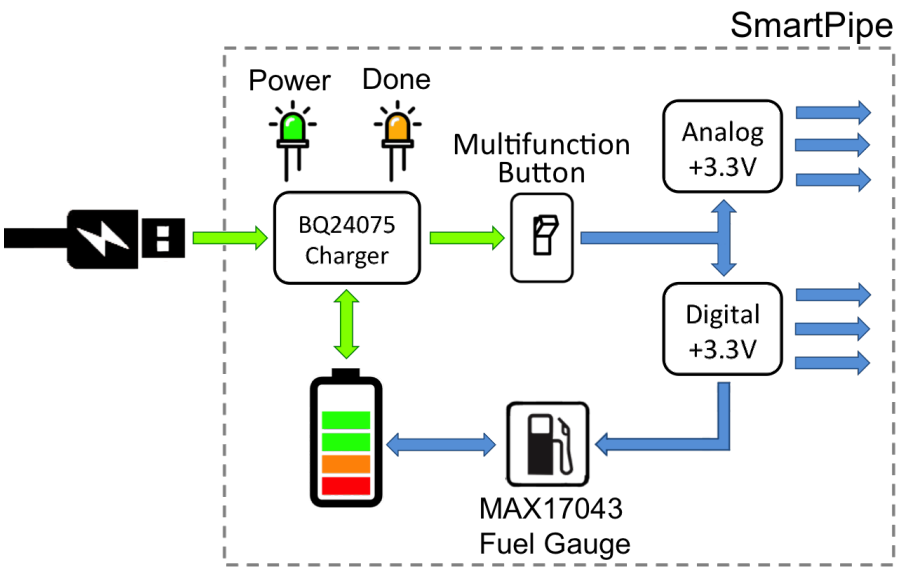


Figure 7.3: Diagram of the SmartPipe's internal power subsystem demonstrating USB power entry and distribution in green, and the systems powered ON by the power switch in blue. The diagram also shows the LiPo battery, its charger and user indication LEDs, and the battery 'fuel gauge' IC

These correspond to the on-board components annotated in Figure 7.4, with the LiPo battery affixed to the bottom of this board with double-sided tape:

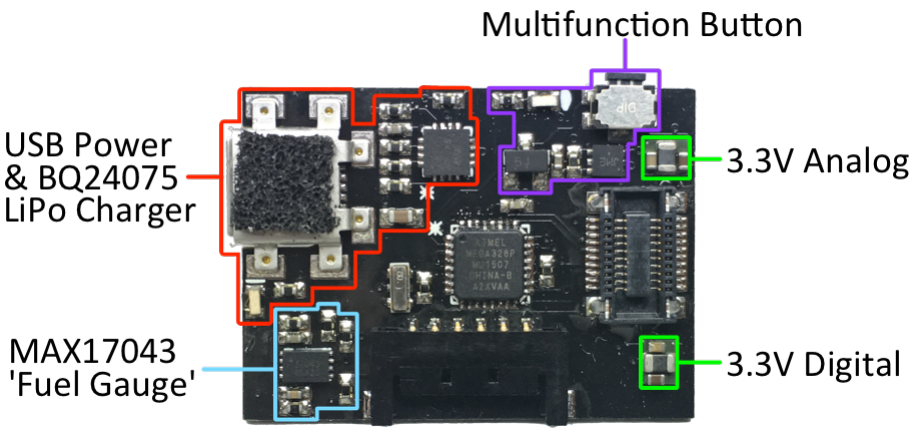


Figure 7.4: The components used to create the power subsystem

The remaining part of this subsection will discuss these in turn.

Power Plane Layout

As mentioned on page 192, the internal power plane was divided into multiple sub-domains in order to route power to different components.

This design is shown in Figure 7.5 below, with the outlines of the various surface

components and their connections to the relevant plane through 300μm through-hole vias.

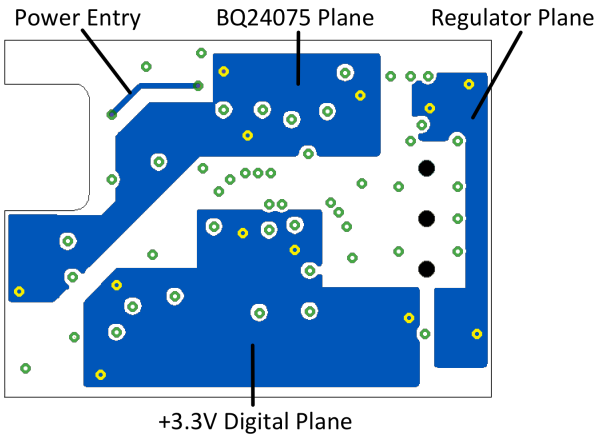


Figure 7.5: Internal layout of the power planes inside the bottom PCB demonstrating the power entry line to the BQ24075 and its plane, the supply to the regulators after the power switch, and the +3.3V digital plane. Vias connecting top layer components to these planes are highlighted in yellow, with other vias in green, and the drill holes to secure the board-to-board connector in black

Note the separation between the plane for the charging circuitry, the plane carrying power from the `ON/OFF` switch to the regulators, and the +3.3V digital plane. The +3.3V analog and digital supplies were carried to the board-to-board connector (page 194) using thick surface traces from each regulator.

Battery Power

Because the SmartPipe is intended to operate as a stand-alone device for $\geq$24hrs or more, I implemented a charging and monitoring system for a high-density lithium-polymer (LiPo) battery. In a final system it may be possible to source a fully-customised LiPo pack to mould into the device, but during this Thesis I had to rely on commercially-available LiPo packs.

The design considerations for this power system therefore had to balance a combination of the total device voltage and current draw (the *power budget*, discussed on page 205) whilst trying to optimise for miniaturisation and longevity.

Charging System

LiPo batteries have complicated timing and magnitude requirements for current and voltage regulation during recharging and discharging phases, and as a result there are a large number of battery power management ICs on the market from multiple manufacturers.

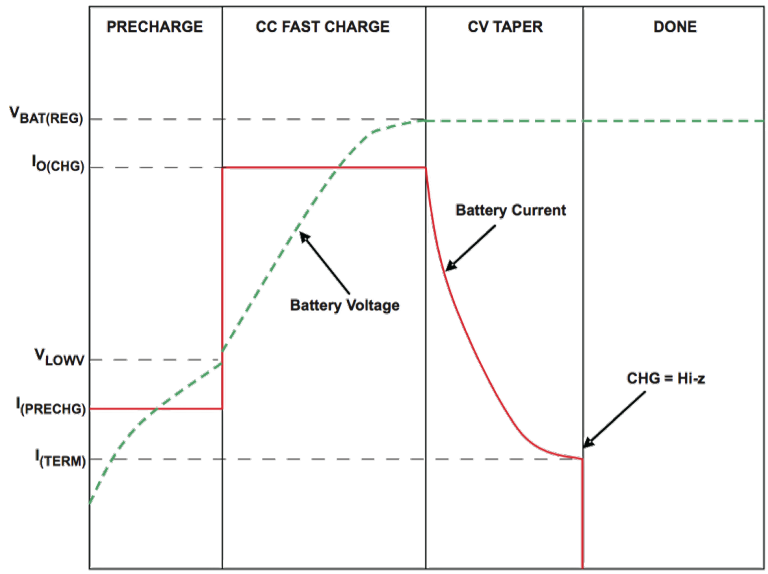


Figure 7.6: The complex current-voltage charging profile of a Lithium Polymer (LiPo) battery. From [125]

I identified the BQ24075 from Texas Instruments as a fully-integrated solution in a small 4mm×4mm 16-pin surface mount (QFN) package matching all of the requirements for the SmartPipe design.

Primarily, it was able to meet the charging profile requirements of a standard LiPo cell of any capacity, with a wide range of customisability. The timings and thresholds were also open to customisation through the use of small external resistors, although the default configuration was sufficient to meet the needs of my design.

The device also had an in-built bypass system to allow for the simultaneous charging of the battery from an external supply and running the system from that supply, a feature which allowed me to use a wall-mounted power supply during prototyping and testing. This often requires additional circuitry in many other charging ICs

I chose to set the BQ24075 to support USB charging compliance (drawing 500mA max) with the `EN1` and `EN2` pins. I also provided user feedback on the supply and charging status via green and orange side-view 1.6mm×1.2mm LEDs mounted beside the USB charging port, connected to the $\overline{\texttt{PGOOD}}$ and $\overline{\texttt{CHG}}$ signals on the BQ24075, respectively.

Battery Monitor

In a battery-powered device it is useful to monitor the remaining charge level of the device not only to provide user with feedback on the likely useful time remaining before the next recharge, but also because LiPo batteries do not tolerate discharge to levels beneath 3V. At voltages below this point the internal resistance increases sufficiently to risk explosions

or incineration if the system is then connected to a recharger. The LiPo pack chosen did contain an under-voltage safety lock-out circuit to render the battery useless if the voltage falls below 2.5V.

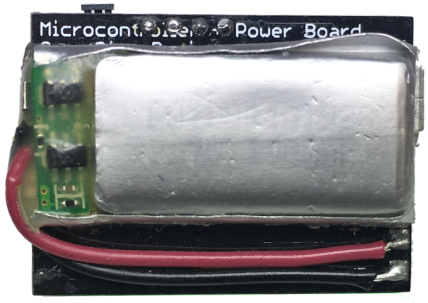


Figure 7.7: Image of the 110mAh LiPo battery mounted on the underside of the bottom digital board of the SmartPipe, demonstrating the pack's integrated under-voltage safety circuit

Unfortunately, neither the BQ24075 nor the LiPo battery packs themselves incorporate any way of monitoring the remaining battery capacity during operation.

For this reason I incorporated a MAX17043[1] 'fuel gauge' IC into the system. This device is an ultra-small (2mm×3mm), 8-pin surface mount chip designed to continuously measure the voltage of single-cell LiPo batteries to ±12.5mV (8-bit) accuracy, and calculate the remaining charge using a proprietary on-chip algorithm.

The MAX17043 communicates directly with the ATMega328P via its I^2C interface to provide the current cell voltage, the % charge remaining, as well as providing a separate `ALERT` interrupt line to inform the host when this charge level falls beneath a threshold programmed by the microcontroller to prevent over-discharge. I set the `ALERT` threshold at 10% residual capacity.

The software library for communicating with the MAX17043 was derived from the open source code published by Luca Dentella in 2012[2] and modified by myself to add the correct cycling of the `ALERT` bit after triggering, and to reduce the unnecessary precision of the charge status to the nearest integer %.

Power Entry

I elected to use a micro-B USB connector for power entry into the board for a combination of small size, availability, low price and robustness of the connection. This only required two of the system's 5 pins (power and ground).

[1] Maxim Integrated, San Jose, California, USA
[2] https://github.com/lucadentella/ArduinoLib_MAX17043

Prior to settling on the ATMega328P, I had attempted to use an ATmega32U4 which had an on-board USB controller for communication with a host computer. However, excessive hardware and software complexity led me to abandon this early design.

USB headers can measure 3mm or more in height. I therefore used a mid-mount USB connector to match the 2mm board-to-board separation. This required a custom mounting slot to be milled from the edge of the PCB, but left the connector only 1.3mm proud of the surface.

DC-DC Voltage Regulators

All components selected for the SmartPipe were capable of operating at +3.3V. Again following the principles of good PCB layout (page 180), I elected to supply the analog and digital +3.3V supplies from two separate DC-DC regulators. Both supplies shared a common ground plane.

The two types of DC-DC voltage regulator in common usage for microelectronic devices are (i) linear regulators (specifically Low Drop-Out (LDO) linear regulators) and (ii) switching regulators. Both of these are intended to supply current to the ICs on the board at a specified voltage, but with different costs and benefits to each.

Linear Regulators

All prototype boards made use of the MIC5205[1] LDO, either in +3.3V or 5V format. These devices were available at very low cost (<£0.50/unit) and able to supply 150mA at a stable +3.3V with 1% accuracy from a supply of 3.45V – 16V, requiring only one 470pF bypass capacitor to stabilise the internal reference, and a 2.2μF capacitor on the output.

While suitable for devices connected to a mains supply such as the DAC (page 183), or with large capacity batteries, I realised that using LDOs in the final design of the SmartPipe would compromise the longevity of the device through various inadequacies including constant power draw and low efficiency.

Figure 7.8 below shows the internal design of a typical linear voltage regulator:

[1]Micrel Semiconductor, now part of Microchip Technology, Chandler, Arizona, USA

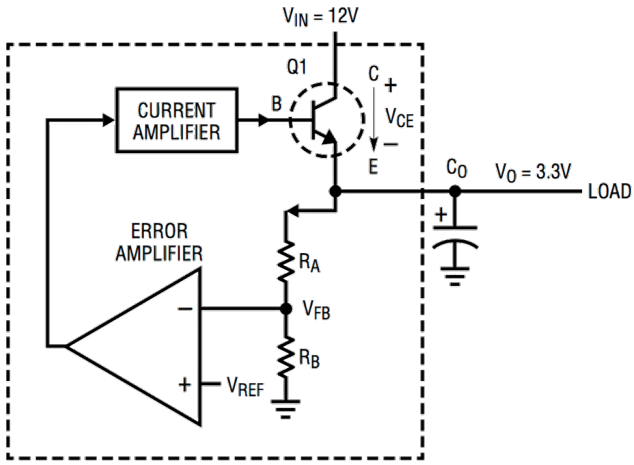


Figure 7.8: The internal design of a linear voltage regulator. From [126]

Linear regulators have a continuous leak current across the internal silicon band-gap reference that feeds into the error amplifier, bias currents at the amplifier's inputs and a continuous current through the gate of the internal FET. These 'quiescent currents' can be as large as a few hundred μA (600μA for the MIC5205 at 100mA), and eat away at a tight power budget in a battery-powered device (page 205).

Furthermore, linear regulators achieve their output voltage by dissipating any voltage headroom between the input and the output target as heat. It is therefore the case that the regulator will sometimes waste more energy as heat than it actually delivers to the device, leading to very low working efficiencies. For example, a +3.3V MIC5205 providing 100mA from a fully charged 4.2V battery has only the following efficiency rating:

$$\begin{aligned}
\text{Efficiency} &= 100\% \times \frac{\mathrm{I_{OUT}} \cdot \mathrm{V_{OUT}}}{(\mathrm{I_{OUT}} + \mathrm{I_{QUIESCENT}}) \cdot \mathrm{V_{IN}}} \\
&= 100\% \times \frac{100\mathrm{mA} \cdot 3.3\mathrm{V}}{(100\mathrm{mA} + 600\mu\mathrm{A}) \cdot 4.2\mathrm{V}} \\
&\approx 78\%
\end{aligned}$$

Meaning that some 22% of the power is continually wasted as heat. At a draw of 100mA at 4.2V before regulating down, this equals as much as 92mW. The numerical efficiency does improve slightly as the LiPo output sags.

I required a better method of providing regulated +3.3V analog and digital supplies on the final SmartPipe to improve its longevity from a small battery.

Switching Regulators

This other class of voltage regulator is able to supply higher currents and stable voltages with much higher efficiencies and in particular, near zero quiescent currents.

To understand why I did not immediately use one of these devices in my system from the outset, it is important to consider the design of switch-mode regulators, shown in Figure 7.9 below:

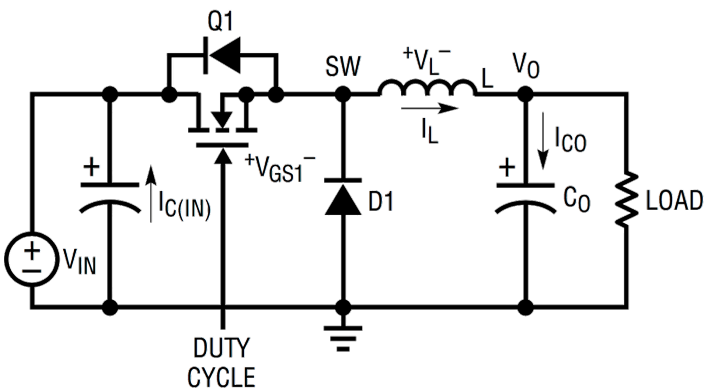


Figure 7.9: Topology of a switching voltage regulator. From [126]

In this design, the switching regulator operates by rapidly turning the current-passing FET on and off at a rate of hundreds of kHz to a few MHz. This transfers aliquots of charge to an inductor's magnetic field in the *on* state, which then collapses and passes to a capacitor in the *off* state giving a DC output voltage with a small ripple. The diode in the circuit provides a conduction path through the inductor while its magnetic field is discharging and the FET 'switch' is open.

The *duty cycle* of the switch determines the rate of charge transfer and therefore the output voltage of the system, and is usually under feedback control from the output. With the FET in the *off* state for most of its duty cycle, the quiescent current draw can be one to two magnitudes less than that of an equivalent LDO, on the order of a few hundred nanoamperes. The efficiency of switching regulators also reaches $\geq$85%.

However, this performance has always come with a few trade-offs, typically including:

- Ripple noise at the output – both the current and voltage at the output can have a ripple. This is countered by large decoupling capacitors with low ESR[1] values.
- Switching noise and EMI – a combination of shielding and/or good layout will resist this noise leaking into sensitive components. Another way of reducing the impact of switching noise is to choose a regulator which switches at frequencies high out of the band of interest.

[1] *Equivalent Series Resistance*

- Larger board surface area requirements – switching regulators typically require more board area than equivalent-output LDOs due to the need for external inductors and capacitors. Higher switching frequencies however reduce the size requirements of these components.
- Higher cost – larger numbers of high quality external components and larger board surface area combine to increase the total cost of implementing a switching regulator over an equivalent-output LDO.

Traditionally the last two issues have confined switch-mode regulators to power adaptors or PCBs with large amounts of board space to accommodate the control circuitry, inductors, diodes and capacitors required.

Fortunately, during the course of this Thesis, Texas Instruments released the TPS82740B, a novel type of switching buck[1] regulator for portable applications which addressed all of these concerns, and which I was able to source from a commercial supplier at a cost of less than <£3/unit.

The TPS82740B device comes in a unique MicroSIP™ surface-mount package which incorporates an inductor, and both input and output capacitors on a single board, with the control electronics implemented on an internal board layer in silicon. It is able to deliver up to 150mA at +3.3V with >90% efficiency, a quiescent current of 360nA, and requires just under 7mm^2 of board area. This device is shown in Figure 7.10, below:

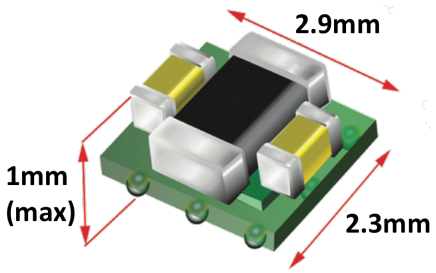


Figure 7.10: Image of the ultra-small-scale TPS82740B fully-integrated switch-mode buck regulator capable of supplying 150mA at +3.3V with >90% efficiency in a 9-point ball-grid array (BGA) package

Two of these devices were used to meet the SmartPipe's power requirements, with one device each for the +3.3V analog and +3.3V digital power lines, at a total cost of <£6. However, I did have to modify my surface mounting techniques in order to accommodate the 9-ball BGA package with a higher solder melting point than that of the paste used to mount the rest of the SmartPipe's components. An image of this component surface mounted on a prototyping board was shown in Figure 5.6 on page 171.

[1]The term 'buck' indicates that this regulator operates to step down the voltage from a higher voltage supply

Multifunction Power Button

The final part of the SmartPipe's power system was the multifunction user interface button under the control of the ATMega328P. The software controlling the button is described further on page 212, but one of its many functions was to act as a soft-latching `ON`/`OFF` switch for the entire device.

Developing this switch by improving on rough designs found on the internet led to a great improvement in the system which had used a physical latching switch to control the power in prototype versions. Using a slide-operated power switch and a small OLED touch display to control functionality, or a rainbow sequence of LEDs to indicate the battery level were either prohibitively costly, power consuming and/or required a lot of processor time to operate.

Most importantly however, a mechanical slide switch would have prevented the microcontroller from being able to force the system to power-down upon receiving a low battery `ALERT` from the MAX17043 'fuel gauge' IC.

The Push-Button Circuit

Instead, a fully-functional interface and soft-latching power switch under control of both the user and the system's ATMega328P was constructed from only one momentary push-button, a dual Schottky diode, a 2mm×2mm dual N/P MOSFET, one resistor and one capacitor. The total solution occupied less than 50mm^2 of board space and results in nanoampere leakage currents through the PMOS in the `OFF` state. The circuit is shown in Figure 7.11 below.

I selected the MCIPTG14K-V horizontal-action momentary switch from Multicomp[1] for its 12V range, audible '*click*' and positive tactile response, as well as its miniature 2.8mm×3.8mm×1.1mm dimensions. Presses were signalled to the user through a single side-view 1.6mm×1.2mm blue LED placed alongside the button and controlled in software.

[1]Premier Farnell plc, Leeds, UK

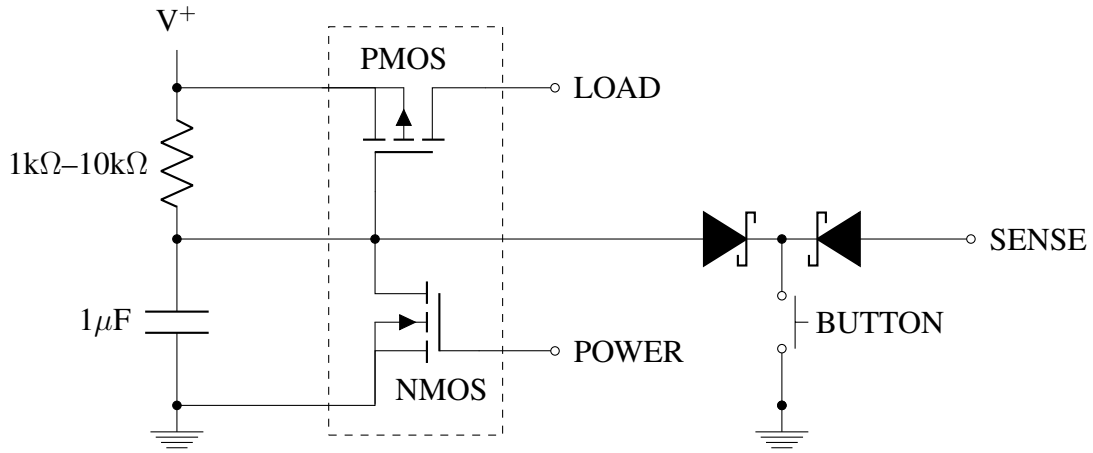


Figure 7.11: Circuit diagram for the multi-function soft-latching power button

- `V+` connects to the output of the BQ24075 battery charging IC
- `LOAD` from the drain of the PMOS connects directly to the power plane that feeds both the analog and digital +3.3V switching regulators.
- `SENSE` connects to an input pin on the ATMega328P that is pulled up by a weak internal resistor (10kΩ–25kΩ)
- `POWER` connects to an output line from the ATMega328P.

Soft-Latching Power, Multifunctional Control

Pressing the `BUTTON` once in the `OFF` pulls the gate of the PMOS to `GROUND` through the Schottky. The RC combination then provides sufficient power to allow the system to boot and begin operating. The first command issued by the microcontroller then pulls the `POWER` line high, activating the NMOS, which keeps the PMOS latched and delivering power to the system from the battery.

Pressing the `BUTTON` when the system is `ON` results in a drop in the `SENSE` voltage to `GROUND` through the Schottky. This controls the behaviour of the system through the button state machine (page 212) which determines whether the user action was a single-click, double-click or a long hold.

Turning the SmartPipe `OFF` again requires a longer 2-second hold of the button and is entirely under the control of the microcontroller in software. The microcontroller pulls the `POWER` connection to `GROUND`, thereby turning off the NMOS and thus PMOS MOSFETs.

The ATMega328P is also able to pull the `POWER` line low in response to the `ALERT` warning from the MAX17043 'fuel gauge' IC, which I selected to trigger at a threshold of 10%.

7.4 Calculating the Power Budget

The power draw of a system during operation is an important consideration when designing any miniature portable device, particularly when trying to maximise longevity on the smallest battery possible.

During the prototyping stages, individual development boards were powered from a bench-top supply or high-capacity 3.7V 440mAh LiPo pack. Later, as components were finalised and the boards reduced in size, a power budget was drawn up which allowed me to reconfigure certain aspects of the circuit and discard or change wasteful components.

For example, I was driven to develop the multifunction power button, and I removed all indicator LEDs from the earlier systems except those driven by mains power (the $\overline{\texttt{PGOOD}}$ and $\overline{\texttt{CHG}}$ signals on the BQ24075 LiPo charger) and left only a single blue LED mounted beside the multifunction switch to provide software-controlled visual feedback on the state of the system.

I also realised that operating the SmartPipe side of the Si8640 digital isolator on the DAC board would have required between 10-20mA of constant current draw, depending on the data rate. For this reason I did not mount this device in the final version of the DAC board (page 185)

Finally, I swapped out the MIC5205 LDOs for the TPS82740B switch-mode regulators which reduced quiescent current draw of the regulators by two orders of magnitude down into the low hundreds of nanoamperes, and improved overall power efficiency by 15-20%.

Table 7.1 below lists most of the components used in the final design and shows the current draws which are mandatory for the normal operation of the system and cannot be reduced any further.

Component	Description	Current	Notes
BQ24075	Battery Charger	50μA	6μA with zero load
MAX17043	‘Gas’ Gauge	75μA (max)	50μA typical
DW01+	LiPo Pack Protector	3μA	
TPS82740	Voltage Regulator $\times$2	360nA	+ 5-10% conversion loss
AFE4490	Oxygenation Sensor	1150μA	
LED	User Interface	1000μA	Used infrequently
Sub-Total		$\approx$1280μA	

Table 7.1: Compulsory current draws by various components

Table 7.2 below shows those remaining components for which the current draw can be

significantly changed in software. Fortunately, these are the most demanding of the system's components and thus changes here can result in the greatest savings to the system's battery life, enabling further miniaturisation of the battery and thus the whole device.

Component	Description	Current	Notes	Page
ATMega328P	Microcontroller	54μA – 45mA	5 power modes	212
DCM03	Oxygen Sensor LEDs	0mA – 50mA		315
LMP90080	Analog Sensor Interface	700μA – 1.7mA	$\propto$ sampling rate	315
LMP90080	Sensor Excitation Currents	0mA – 2mA		315
Sub-Total		750μA – 100mA		

Table 7.2: Current draws that can be modified through changes in software

After implementing changes in software to minimise the current requirements of the final system (detailed later on the pages listed in Table 7.2), the system was able to operate continuously for nearly 18 hours from just a small (15mm$\times$25mm$\times$4mm), low-cost (<£10), 110mAh 3.7V LiPo battery.

7.5 Designing the SmartPipe's Software

The ATMega328P is a single-threaded 8-bit 16MHz processor with 32KB of program memory and 2KB of RAM. With such limited processing power and memory, it is not capable of supporting a real-time fully multitasking operating system.

However, the ATMega328P only needs to manage a small range of functionality to operate the SmartPipe. These are:

1. Reading sensor data from the ICs on the middle mixed-signal PCB
2. Relaying this data to the DAC
3. Shutting down if the battery voltage drops dangerously low
4. Responding to infrequent user input on a single button

Whilst the first three behaviours should be turnkey and automatic, user input was necessary for the following:

- Powering the system `ON` and `OFF`
- Placing the blood pressure sensor into `CALIBRATION` mode
- Capturing the `CALIBRATION`
- Requesting visual indication of the current state of charge of the battery

The details of the circuit to implement the multifunction button were discussed previously on page 203. The rest of this Chapter will discuss the software I designed and implemented

based on three parallel interrupt-driven state machines, in order to run the SmartPipe whilst consuming as little battery power as possible.

This software was developed in parallel with the selection and testing of new components and designing prototype boards. As a result, it took nearly 18 months of part-time development work in total to reach the final version, producing 1 master program calling upon 6 code libraries, not to mention the many others that were discarded as development shifted focus. The end result was a streamlined, fully-functional system implemented in ≈2,500 lines of code, shown below in Figure 7.12.

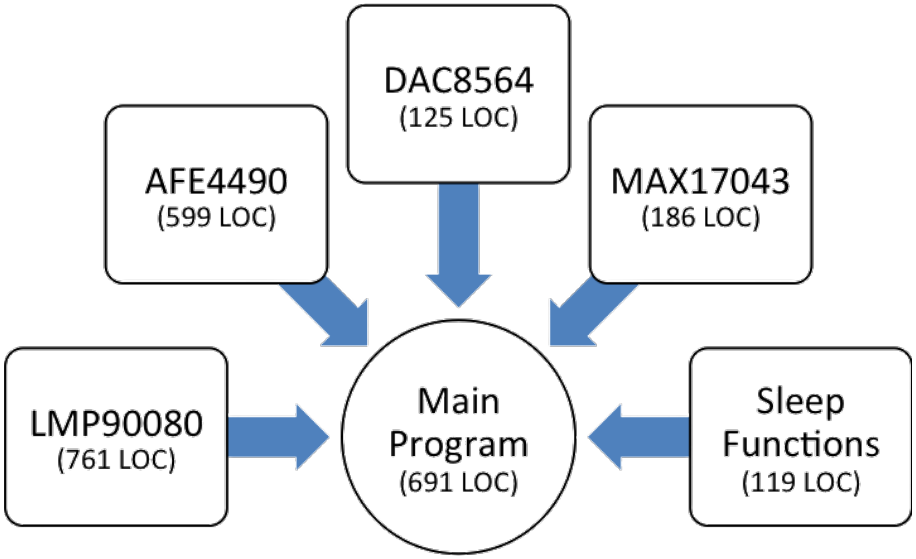


Figure 7.12: The software libraries developed to implement the SmartPipe, including the lines of code (LOC) for each

Interrupts

An interrupt is a trigger signal to alert the microcontroller that a change has taken place in the outside world. This can mean the completion of a conversion cycle by an ADC, the push of a button by a user, a signal to indicate that a timer has finished, or an alarm that something is going wrong.

The primary advantage of using an interrupt-driven system is that the microcontroller only handles events as and when they happen. It can therefore be kept in a low power state, ignoring the outside world, until its attention is actually required.

The alternative would be to continuously poll an input or a register to watch for transitions. The polling rate then determines the potential maximum response rate of the system to external events and greatly increases its power consumption.

Whereas sophisticated microcontrollers often have each pin addressable by its own interrupt

flag, and user-selectable interrupt priorities, the ATMega328P only has a small pool of interrupts, and only one interrupt priority setting. This means that sequential interrupts will be ignored until the first interrupt has been handled and the master flag reset.

These interrupts are largely self-explanatory:

- Brownouts – for alerting the system that a system check may be necessary upon reset
- Timer overflows – for running internal clocks and watchdog timers
- Serial port activity – the UART, SPI and I^2C systems are all interrupt-driven
- ADC complete – conversion completion by the internal 10-bit ADC
- Voltage change on an external digital pin

The ATMega328P has 5 pin change interrupt vectors. The first two of these, `INT0` and `INT1` are hard-wired directly to digital pins `PD2` and `PD3`. These particular interrupts are able to detect not only voltage changes from `LOW` to `HIGH` and vice versa, but also `RISING` or `FALLING` edges.

This is particularly useful for external devices chosen for the mixed-signal PCB such as the LMP90080 or AFE4490 (discussed in Chapter 9) that signal the completion of their conversions through `FALLING` and `RISING` edges on their `DRDY` pins, respectively.

The 3 remaining pin change vectors (`PCINT0` – `PCINT2`) only trigger on absolute changes (`LOW` to `HIGH` or vice versa) rather than the direction of the change. Furthermore, each vector serves a bank of pins:

`PCINT0` serves pins `PB0` – `PB7` (digital pins 0 – 7)
`PCINT1` serves pins `PC0` – `PC6` (digital pins 8 – 13)
`PCINT2` serves pins `PD0` – `PC6` (digital pins 14 – 19 $\equiv$ analog pins 0 – 5)

Therefore, a voltage change on either `PB0` or `PB1` will trigger `PCINT0`, and the microcontroller will then have to determine which of the two triggered the interrupt and respond accordingly.

Interrupt Service Routines

The functions that handle the response of the microcontroller to the triggering of an interrupt are termed *interrupt service routines* (ISRs). Because the ATMega328P lacks the ability to assign interrupt priorities or create nested interrupts, the ISRs should be kept as short as possible because the system will be blind to all other events until completion.

This is straightforward in the case of handling the triggering of `INT0` or `INT1`. Here is my code for setting up and handling `INT0`, which listens for a `RISING` level change on the `DRDY` from the AFE4490 IC on the middle PCB:

```
#define AFE_DRDY  2 //AFE DRDY feeds into INT0 on digital pin 2

```

```
void setup()
{
pinMode(AFE_DRDY, INPUT)
attachInterrupt(digitalPinToInterrupt(DRDY_AFE),AFE_Interrupt,RISING);
}

void AFE_Interrupt(void)
{
  newAFE = true;
  Sleep.wakeupSPI();
}
```

Here the triggering of `INT0` by a rising edge informs the main loop that there is new data ready for collection from the AFE and wakes the SPI module from sleep in preparation for communication. This ISR takes approximately 10 instructions in machine code, or 625 nanoseconds to complete.

The handling of the `PCINT0`, `PCINT1` or `PCINT2` vectors however require the recall of previous pin states from memory with sequential comparisons and branching which can add tens of μsec to the duration of the ISR.

This directly impacts upon the maximum sampling rate of the system once added to the SPI communication overheads and other branching instructions in the main loop, and in the worst case it could lead to data loss from other interrupt-driven services.

Blocking and Non-Blocking Code

An ISR is a form of *blocking* code. This indicates that all other functionality is 'blocked' from executing until that section of code completes. Large portions of blocking code such as timer-based LED blinking routines and counting loops can interrupt the expected behaviour of the system, potential leading to data loss or 'locking-up'.

The code for the SmartPipe made exclusive use of non-blocking code for its timers and state transition logic. The only blocking routines were the ISRs themselves.

State Machines

A state machine is a way of controlling program flow that allows the designer to regulate the transition mechanics between one functionality and another. This prevents 'free-wheeling' and 'non-determinacy', where the system may require otherwise complex comparative or

branching logic to handle multiple concurrent events so as to avoid becoming trapped in an unexpected pattern of behaviour.

Instead, the conditions and the behaviour of state machines never deviate from a rigid path. They also provide a robust method for handling errors outside the expected range of system behaviour by falling back on default states.

7.6 Main Program Loop

The Arduino IDE is designed to implement software on the ATMega328P as two main functions, (i) the initialisation or `setup()` function which runs once at startup, and (ii) the main program `loop()` which runs continuously so long as the system is supplied with power.

The very first task for the `setup()` function was to drive the `POWER` line `HIGH` to keep the system `ON`, because the 10msec lifetime of the RC 'battery' implemented in the multifunction switch circuit (Figure 7.11, page 204) was only sufficient to initiate startup.

The rest of the `setup()` was used to initialise variables, define the pin-map of the ATMega328P for interacting with its peripheral devices, and to send initialisation signals and code packets to the various peripherals over the appropriate communication protocol (I^2C or SPI). The particular order of the various initialisation routines did particularly matter, except that the master ISR flag was held uninitialised until the end of startup so that the ISRs couldn't trigger before completion.

The system then indicates completion of the startup process by flashing its blue LED twice using a non-blocking, timer-based loop.

The main `loop()` had to be organised more carefully.

Referring back to Table 7.2 on page 206, the power draw of the ATMega328P can vary from 54μA in deep sleep, all the way to 45mA in normal operation when communicating with multiple peripherals and performing mathematical operations.

The only way to reduce the current draw by the microcontroller and achieve a $>$24hr functional lifetime from the SmartPipe was to reduce the time spent in the normal operating mode as much as possible, and instead remain in one of the power-saving sleep modes until triggered to wake up by an interrupt.

I wrote the main program `loop()` as follows:

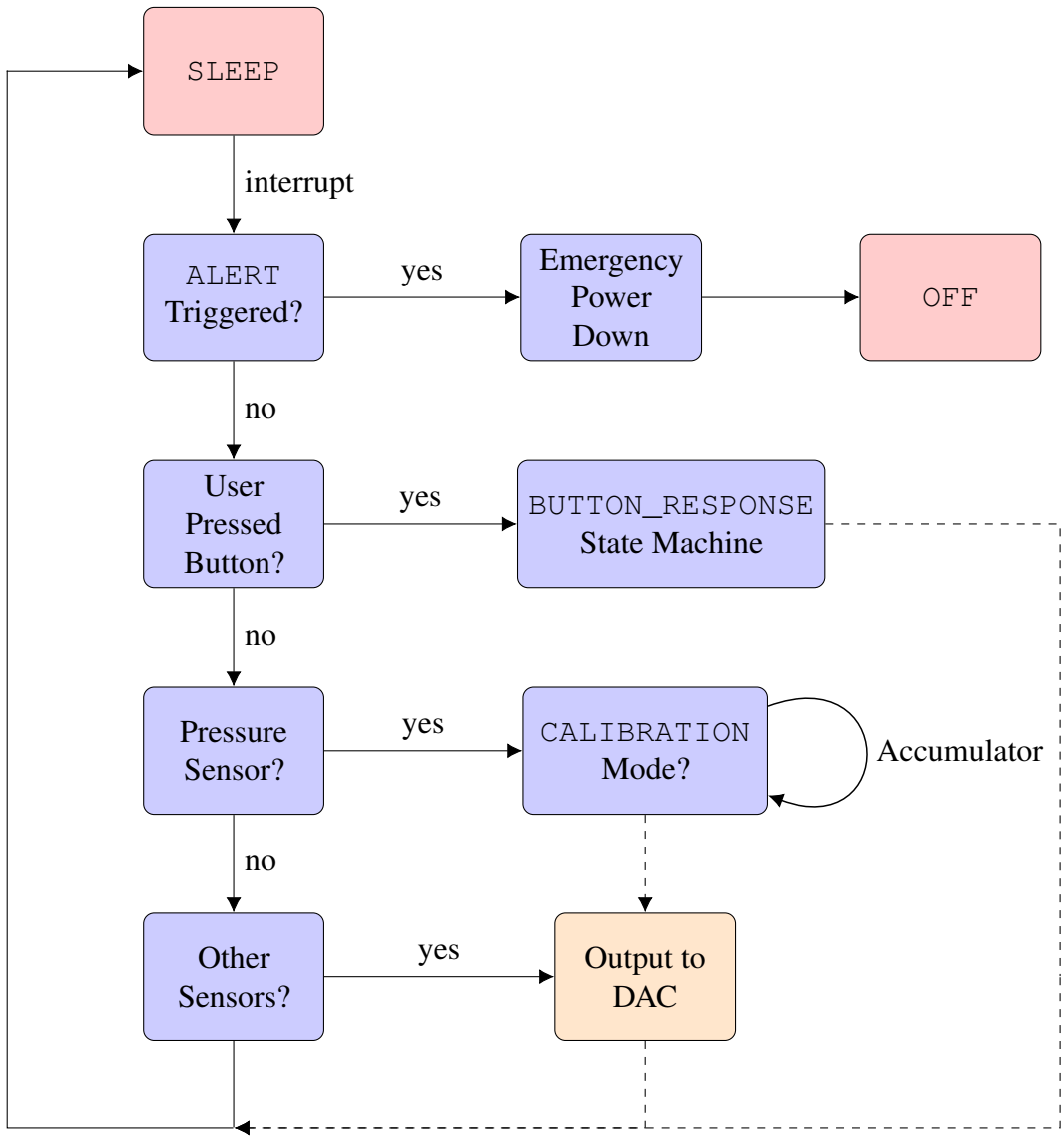


Figure 7.13: Diagram of the main program loop. The loop terminates with the system returning to `SLEEP`. The detail of the `CALIBRATION` mode is described later on page 214

The ISRs were written only to set a flag to indicate to the master `loop()` that the interrupt had fired, minimising the time in the ISR and therefore blocking. The ATMega328P automatically wakes from sleep in response to interrupts, to the point in code just after it was sent to sleep.

The system then rapidly assesses which ISR was triggered according to the flags set, and branches into that pathway. Upon completion and returning to the main loop, the program quickly returns the microcontroller to sleep.

Power Savings in Software

The ATMega328P should spend the majority of its time in the `STANDBY` power-saving mode. Unlike the most power-saving `POWER_DOWN` mode, `STANDBY` keeps the external oscillator active, which decreases the response time to the multiple frequent interrupts it was expected to receive. I disabled the following microcontroller functions in software in order to save as much power as possible:

- Brownout detection – requires an internal current leak through a voltage reference
- Analog input buffers
- The analog comparator
- The internal ADC
- USART controller
- Timers 0, 1 and 2
- I^2C and SPI hardware - powered up if the relevant interrupt fires

These changes in software reduced the current drawn by the ATMega328P from 8.1mA to 3.9mA, a saving of 4.2mA.

Again referring back to Table 7.2 on page 206, additional power savings were achieved by adjusting the internal timers and other features of the sensor-interfacing peripherals. The details of these changes will be covered later on the pages listed in that Table.

Implementing the User Interface

The user interface was managed through the interaction of three state machines running in parallel, called `BUTTON_STATE`, `BUTTON_RESPONSE` and `SYSTEM_STATE`.

Button Response Handler

The button was debounced in hardware due to the effective RC filter at the soft-latching button. Each press was programmed to elicit a brief flash from the LED to provide user feedback.

The handling of the physical state of the button was controlled by an ISR linked to `PCINT1` (page 208), triggered by any level change on the input pin. This requires the ISR to compare the level against the previous level in order to determine whether the button had been `PRESSED` or `RELEASED`. The transition back to `WAITING` was handled in the `BUTTON_RESPONSE` code.

Constraining the transitions with a state machine allowed the system to determine whether the user had executed a single-click, double-click or long hold through the use of non-

blocking timers. It also meant that it would be straightforward to extend the functionality of the single button as required, as well as defining impossible transitions.

The durations of the delays were hand-tuned after testing by my colleagues.

Figure 7.14 below demonstrates the operation of the BUTTON_RESPONSE state machine for determining user action:

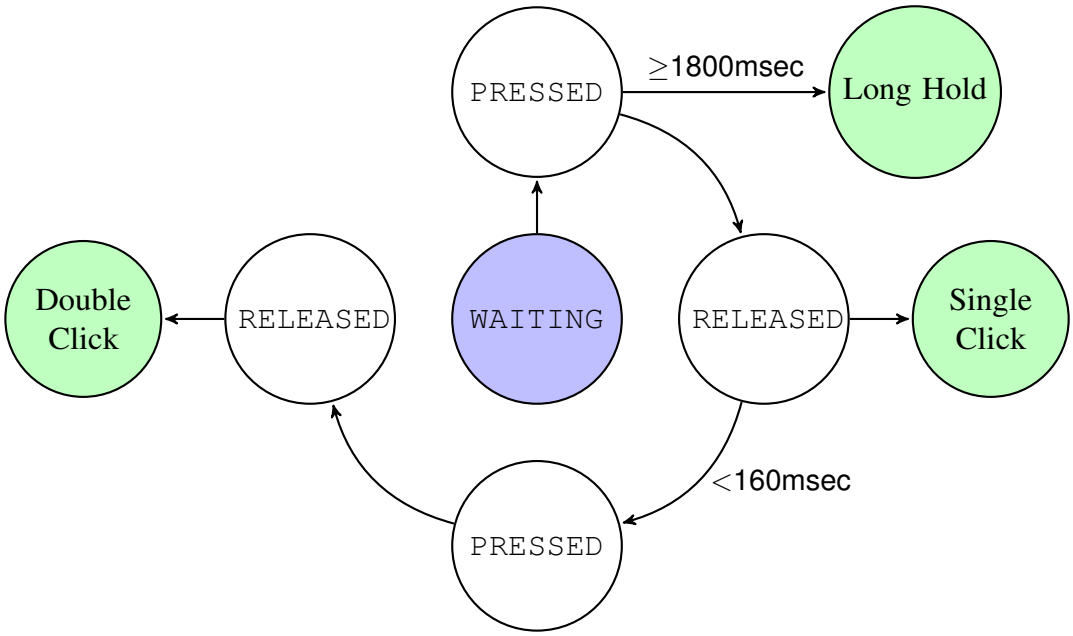


Figure 7.14: Diagram of the transitions in the BUTTON_RESPONSE state machine

System States

The behaviour of the BUTTON_RESPONSE state machine in Figure 7.14 fed into the SYSTEM_STATE machine shown below in Figure 7.15 in order to control the transitions between different SmartPipe functionality:

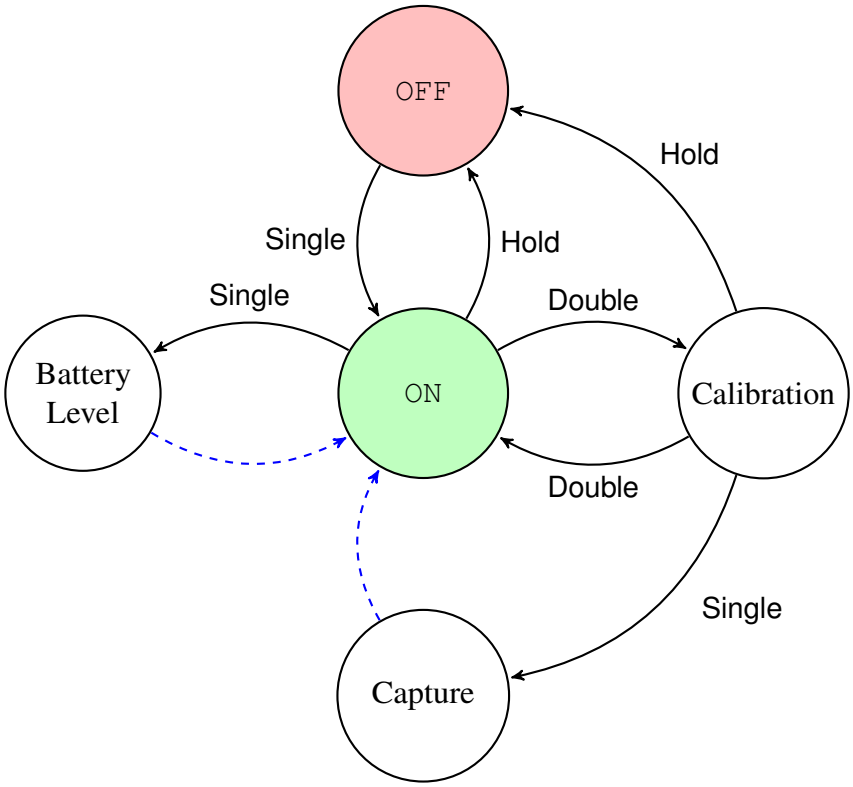


Figure 7.15: Transitions of the `SYSTEM_STATE` state machine demonstrating the button presses necessary to transition to and from `CALIBRATION`, indicating the battery level or the interrupt-driven main loop (`ON`). The blue dashed lines indicate automatic state transitions

After returning to the `ON` state, the system is quickly returned to sleep by the main `loop()` to await external interrupts. The transition back this state resets the `BUTTON_STATE` back to `WAITING`.

Calibration Mode

The user is able to enter the `CALIBRATION` state by double-clicking the button from the `ON` state. Here, input from the blood pressure sensor is continuously fed into DAC as well as a 16-sample rolling accumulator as an averaging filter, until the user 'captures' the data by using a single-click on the button, or quits without calibration through another double-click. This value is then used to provide a calibrated and scaled pressure output using the 5V full scale of the DAC and PowerLab, as discussed previously on page 184.

The `CALIBRATION` state expects the SmartPipe to be connected to a calibrated sphygmomanometer at 180mmHg of pressure for correct scaling. Calibration of the sensor in this manner is covered later on page 300 in Chapter 9.

The blue LED provides the user with feedback in the form of a repeating double-flash with non-blocking code so long as they remain in the `CALIBRATION` state.

Battery Level Indicator

Also under user control is the battery level indicator. A single-click during normal operation will cause the system to read the charge level from the MAX17043 and display the result on the single blue LED as a number of flashes, where each flash corresponds to the following 25% brackets:

1 Flash:	`ALERT` level → 26%
2 Flashes:	27% → 51%
3 Flashes:	52% → 75%
4 Flashes:	76% → Full

Power Off

Finally, the user is able to power the system off with a long hold of the button. Referring back to Figure 7.11 on page 204, this action pulls the `POWER` pin to ground, thereby turning off the FET and detaching the voltage regulators and the rest of the system from the BQ24075 and battery.

Because this transition is handled entirely in software, the ATMega328P is also able to drive the system into the `OFF` state in response to a low-battery `ALERT` from the MAX17043 'fuel gauge' IC.

In the `OFF` state, the only power draw is the small leakage current from the battery's internal electronics and BQ24075 charger in power-down mode as shown in Table 7.1 on page 205. This is around 10μA – 12μA in total, meaning that the 10% `ALERT` setting in the MAX17043 gives the user 6 weeks to recharge the system before risking battery damage.

Turning the system `OFF` results in a succession of 10 quick flashes from the blue LED. The shutdown process cannot be halted once started and so this sequence uses blocking code.

Chapter 8

The Top (Sensor) Layer of the SmartPipe

This Chapter discusses the decisions that went into selecting the temperature, flow, pressure and oxygen sensor systems for the SmartPipe, as well as designing the pipe into which they were integrated.

8.1 Decisions Regarding the Sensor Suite

The discussion of the design requirements for the sensor suite were initially presented in Chapter 5. Table 5.1 from page 166 has been reproduced below for ease of reference.

	Temperature	Pressure	Flow	Oxygenation
Range	0 – 40°C	0 – 200 mmHg	0 – 600 ml/min	75% – 99.9%
Sensitivity	0.1°C	2 mmHg	10 ml/min	1%
Bandwidth	1 Hz	200 Hz	2 Hz	2 Hz

Summary of the design requirements for the SmartPipe's sensor system

These requirements should be achievable through careful component selection and circuit design. The rest of this Chapter will discuss those steps in turn for the temperature, flow, pressure and oxygenation sensor systems, followed by the design of the 3D pipe into which they have been integrated.

The next Chapter will discuss the selection of the integrated circuits to interface with these sensors and show the results of testing.

8.2 Selecting a Temperature Sensor

In its fundamental form, temperature is a representation of the average kinetic energy of atoms or molecules within a substance, exploited since the 1700s in the form of the bulb thermometer. These translate the volume expansion of a liquid as the kinetic energy of its atoms increases into the movement of a column of liquid within a fine tube of constant diameter

A problem arises when we wish to translate the analog property of mean particle kinetic energy into an electronic signal for recording, manipulation and interpretation by a digital system.

Two main methodologies I considered for the electronic detection and measurement of temperature were:

(i) The thermoelectric effect exploited by thermocouple devices
(ii) Thermistors such as platinum resistance temperature detectors (RTDs)

The Thermoelectric Effect

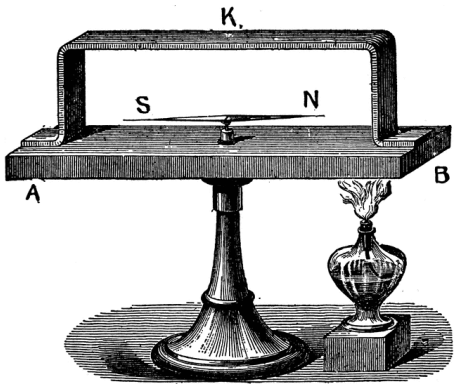


Figure 8.1: Image of Seebeck's original experimental design demonstrating the deflection of a compass needle (N and S) by heating a junction between bars of antimony and bismuth (the platform A–B, and K). ©2016 *ClipArt ETC*, The Florida Center for Instructional Technology

The thermoelectric effect is believed result from charge carriers moving at different relative rates across the points of contact between two different conductive materials. High energy electrons in the material with lower binding energy (higher conductivity) will be liberated from the warmer junction and migrate to the point of least energy, or 'cold' junction.

This migration alters the charge density at each junction, thereby creating a measurable potential difference between the two. If allowed to remain in contact for infinite time, the

potential will eventually fall to zero as the charge carriers redistribute evenly throughout the material.

By applying a known temperature gradient across a material and measuring the induced potential difference, we may derive a constant of proportionality known as the *relative* Seebeck coefficient in μVolts per Kelvin. The term *relative* is used because the act of measuring the potential difference requires one to apply probes made of a conductive material which is different to that of the material under investigation, thereby creating additional junctions with their own potential differences. Furthermore, the coefficient itself is known to vary with temperature in many materials.

Thermocouple Devices

A thermocouple probe is the junction of two conductive materials (or semiconductors) with known Seebeck coefficients. The potential difference between the ends of the sensor will be proportionate to the temperature of the junction, and this signal can be digitised.

The materials used in the construction of the thermocouple are cheap and readily available, and the resulting sensors function without any excitation source. Table 8.1 below provides a list of the most commonly used thermocouple materials and their sensitivities.

Type (ANSI)	Positive Conductor	Negative Conductor	Temperature Range (°C)	Error (± °C)	Sensitivity at 20°C (μV/°C)
T	Copper	Constantan	-200 to +350	1.0	40
J	Iron	Constantan	0 to +750	2.2	51
E	Chromel	Constantan	-200 to +900	1.0	62
K	Chromel	Alumel	-200 to +1250	2.2	41
S	Pt (10% Rhodium)	Rhodium	0 to +1450	1.5	7

Table 8.1: Table of common thermocouple materials demonstrating their sensitivities defined at 20°C, and full-scale error range. Constantan and Alumel are specialised Cu-Ni and Al-Ni alloys, respectively

Problems with Thermocouple Devices

Unfortunately thermocouple sensors are not without problems, some of which make them unsuitable for the SmartPipe concept.

- Low sensitivities in the range of tens of μV/°C
- Non-linearity of the Seebeck coefficient requiring lookup tables or polynomial fitting
- Need for cold junction compensation with external hardware
- Large probe sizes not easily miniaturised or incorporated into the SmartPipe

These will be explored briefly.

Non-Linearity

Relative Seebeck coefficients are non-linear, resulting in voltage outputs that are non-linear over temperature. Figure 8.2 below illustrates this for a number of different common thermocouple materials. Even K-type thermocouples, considered the most linear of the materials, has a non-linear Seebeck coefficient over the biological temperature range of 0 – 40°Celsius.

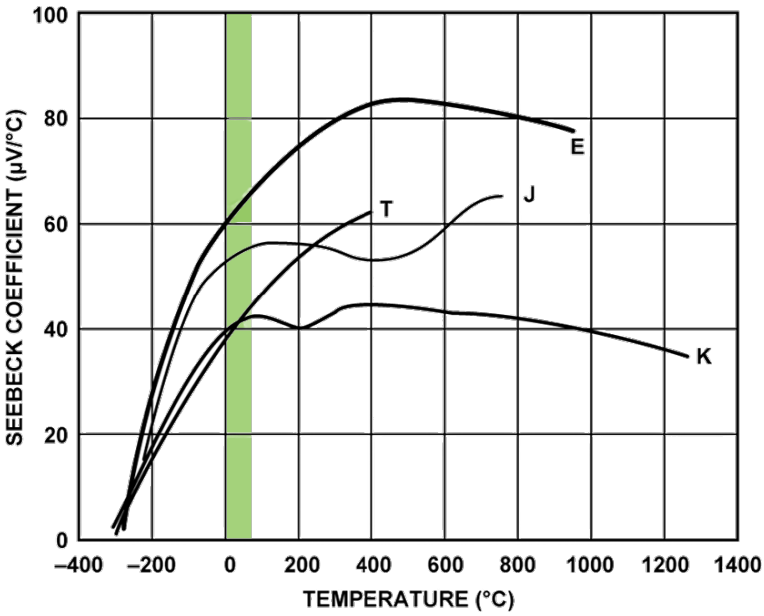


Figure 8.2: Non-linearity of the Seebeck effect over temperature for common thermocouple materials. The biologically-relevant temperature range of 0 – 40°C is highlighted in green. Image from [127]

Cold Junction Compensation

Thermocouples only generate potential differences when they form a complete thermoelectric circuit, comprising a 'hot' junction at the working end of the sensor, and a 'cold' junction across which the potential difference is measured.

The potential difference that develops is a function of the difference between both of these temperatures, and so the circuit can be considered to be thermally 'floating' without a fixed reference point. Note that the act of measuring the potential difference creates further junctions between the different conductive materials used for the sensor and the connection to the circuit board which may be a led/tin solder in contact with copper.

One solution to the multi-junction problem has been to hold another thermocouple's 'cold' junction at 0°C in a bath of ice-water and to use this as a fixed and known reference point for the other thermocouple circuit under measurement, to which it connects in series. In

fact, the NIST's thermocouple tables have been drawn up assuming this compensation and referencing methodology.

Two solutions have been developed to obviate the need for a complicated and bulky ice-bath apparatus. Both rely on the use of an *isothermal block*.

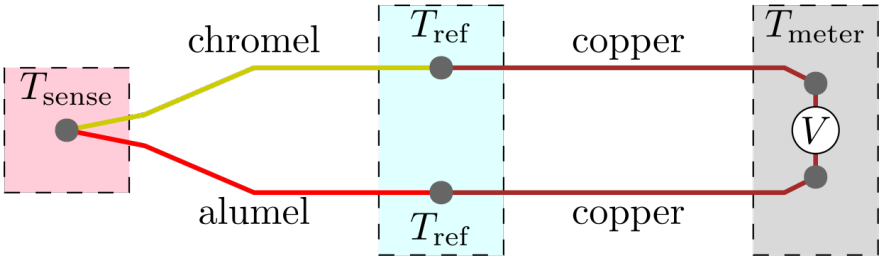


Figure 8.3: Circuit for a K-type thermocouple (chromel/alumel) demonstrating the use of an isothermal block in light blue to stabilise the reference voltages T_{ref} and read those values using identical copper wires. This does not show the second sensor required to read the temperature of the block itself

(i) Software compensation

In Figure 8.3 above, both of the 'cold' junctions are connected to another known material, such as copper. By measuring the temperature of this block with a separate sensor, this provides a known reference point against which the temperature of the thermistor's 'hot' junction can be determined.

(ii) Hardware compensation

This is sometimes termed an 'electronic ice point' compensation. Here, a DC voltage is applied across the 'cold' junction in proportion to the temperature of the*isothermal block* so that the 'cold' junction is a summation of the 'electronic ice point' voltage and the voltage generated by the Seebeck effect.

Probe Size and Manufacturability

Even with the comparatively good linearity and sensitivity of E-type thermocouples, deploying a thermocouple in my design was limited by the availability of these devices in miniature form-factors.

Encapsulated micro-thermocouples are available in wire diameters of 36AWG – 44AWG (127μm – 50μm diameter) but are not available in surface-mount configuration and must therefore comprise a body protruding into the flow. In terms of microengineering and manufacturability of the connections to the 'cold' junction, the connecting wires of these probes may prove too difficult to attach to the rest of the circuitry by hand soldering.

Not only that, but the compensation of the 'cold' junction itself requires additional sensors and electronics to be developed and deployed.

Summary of Problems

Despite being cheap and robust, with rapid response times and wide temperature ranges, problems with poor linearity, low sensitivity, complicated compensation methodologies and poor miniaturisability without specialist equipment meant that thermocouples did not make an attractive proposition for temperature sensing in the SmartPipe concept.

We will now discuss an alternative sensor methodology which is more suitable for use as a biological temperature probe.

Platinum Resistance Temperature Detectors (RTDs)

Certain pure metals, such as platinum, exhibit an increase in their bulk resistance with temperature due to the increasing disorder of the material on an atomic scale. With increasing mean atomic kinetic energy, electrons will collide and exchange energy more readily, decreasing their ability to flow through the material, and thus increasing the resistance to current flow. Devices derived from pure metals are known as *Resistance Temperature Detectors*, or RTDs.

Modern metallurgical techniques allow for metals of precise composition to be drawn into ultra-fine filaments, or electrodeposited onto an insulating substrate to create a conductive trace.

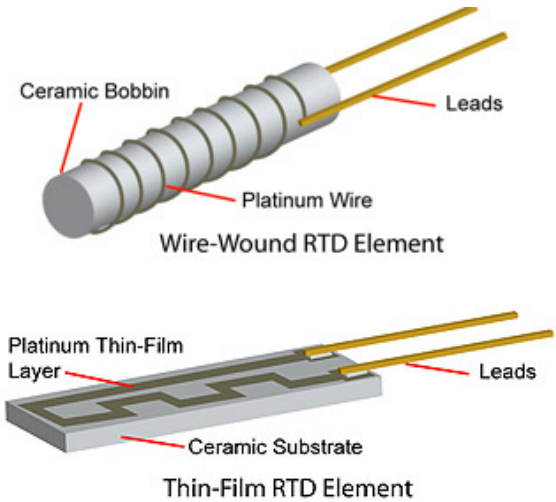


Figure 8.4: Illustration of a wire-wound vs. thin-film RTD. Surface-mount RTDs are an extension of the thin-film type, an example of which is shown in Figure 8.6 on page 225

RTDs are commonly manufactured from high purity platinum, with known thermal characteristics referenced and standardised at 0° Celsius, so-called $R_{\emptyset}$. Platinum RTDs are most commonly available with $R_{\emptyset}$ values of 100Ω or 1000Ω.

Advantages

There are multiple benefits to choosing a platinum RTD over a thermocouple, including improved (i) sensitivity, (ii) linearity, (iii) accuracy, and (iv) ease of incorporation into the SmartPipe design.

Sensitivity

The sensitivity of a platinum RTD is given by a characteristic known as *alpha* (α), which describes its change in resistance per degree celsius (Ω/Ω/°C) over the range 0 – 100 °C. The α values for commonly-used RTD materials are given in Table 8.2 below:

Metal		Temperature Range (°C)	α (Ω/Ω/°C)
Platinum	Pt	-240°C to +660°C	0.003926 (≥99.999% pure Pt) 0.003911 (US IPTS-68) 0.003850 (IEC 60751) 0.003750 (Low-cost Pt RTD)
Copper	Cu	-200°C to +260°C	0.00427
Nickel	Ni	-80°C to +260°C	0.00617
Molybdenum	Mo	-200°C to +200°C	0.00300

Table 8.2: α values for common RTD materials. Note the multiple standards of platinum available

A typical IEC 60751-compliant platinum RTD with an α of 0.00385Ω/Ω/°C (or 3850ppm/K) and an $R_{\emptyset}$ value of 100Ω will increase in resistance by 0.385Ω/°C.

Note that other materials have higher α values than platinum, including cheaper metals such as copper or nickel. This does give them greater sensitivity, but with other disadvantages such as poorer biocompatibility.

Linearity

Platinum RTDs are highly linear devices, particularly over the temperature range of 0°C – 40°C. This becomes evident in Figure 8.5 below, illustrating a head-to-head comparison of

relative linearity against a K-type thermocouple[1]. These values were derived by taking the % deviation from an idealised linear response from known values at 0°C and 40°C.

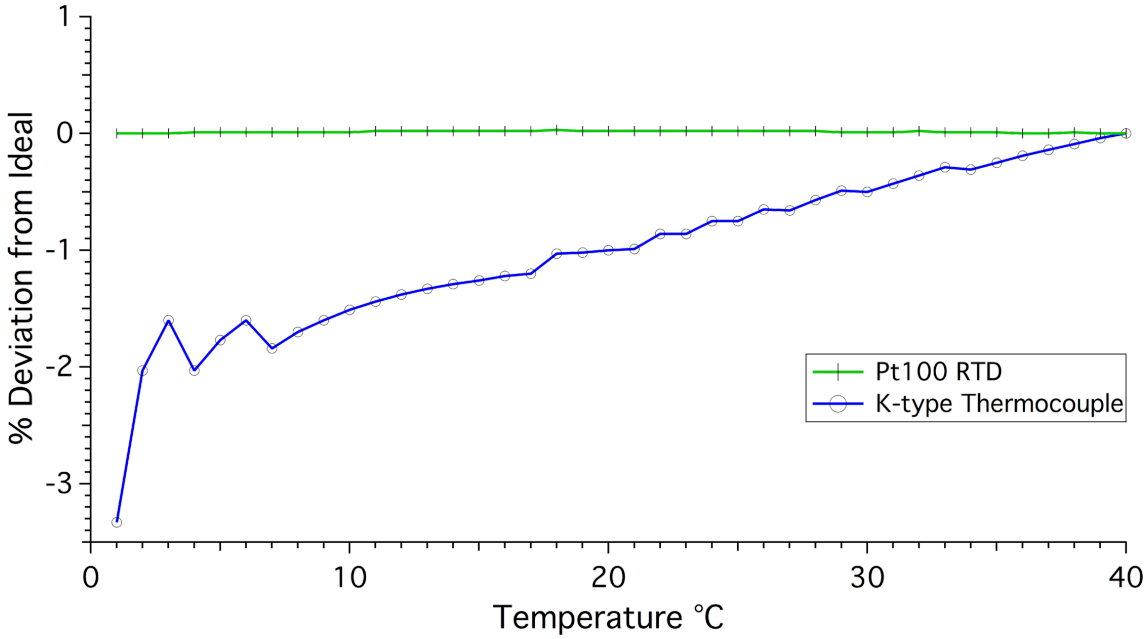


Figure 8.5: Comparison of the nonlinearity of a K-type thermocouple and a Class A IEC 60751 Pt100 RTD as the % deviation from a straight line through the idealised response for calibrated values between 0°C and 40°C [128]. Maximum step-wise deviation for the RTD is <0.1%

There are also international standards defining the linearity and tolerance limits of commercially-available RTDs. For the typical 3850ppm/K platinum RTD, IEC 60751 defines the acceptable performance characteristics for Class A and Class B sensors as follows:

Characteristic	IEC 60751 Class A	IEC 60751 Class B
$R_{\emptyset}$ (Ω)	$\pm$ 0.06%	$\pm$ 0.12%
α (Ω/Ω/°C)	3850ppm $\pm$63ppm	3850ppm $\pm$63ppm
Range (°C)	-200°C to +650°C	-200°C to +850°C
Tolerance ($\pm$°C)	0.15 + 0.002* |T|	0.30 + 0.005* |T|

Table 8.3: Characteristics of Class A and B Pt100 RTDs according to IEC 60751

However, the value of α is not perfectly constant, and does change slightly over the entire temperature range.

[1] K-type thermocouples are considered to be the most linear type

For this reason an additional fully-characterised value is given with every α, known as δ. Where α gives the absolute change in resistance over a defined range, δ describes the change in α over the same range.

This nonlinearity was taken into account in the equations of H.L. Callendar (1863 – 1930), described further on page 227.

Accuracy

The tight tolerance standards of RTDs shown in Table 8.3 demonstrate the high degree of accuracy of Class A and B sensors.

For example, the response of a Class A sensor with a tolerance of (± 0.15°C $+ 0.002 * |$T$|$) will be constrained to within just $\pm$0.23°C across the entire biological range of 0°C – 40°C.

RTD sensors are also very thermostable, typically in the range of 0.1 °C per year of rated operation. When combined with the high linearity and accuracy of these sensors, it is clear that they far outperform thermocouples over the range required for the SmartPipe.

Incorporating the Probe into the SmartPipe

Platinum RTD probes are marginally more expensive than thermocouples, but are available in miniature chip packages that are more readily manageable for laboratory manufacturing using standard surface mount techniques.

I chose to use a Class A Pt100 RTD which is available in 1206 SMD format from IST GmBH[1], at a cost of less than £8 per unit[2].

Such a package configuration matches the SmartPipe concept perfectly, because it can reside in the wall of the pipe and not interfere with the flow.

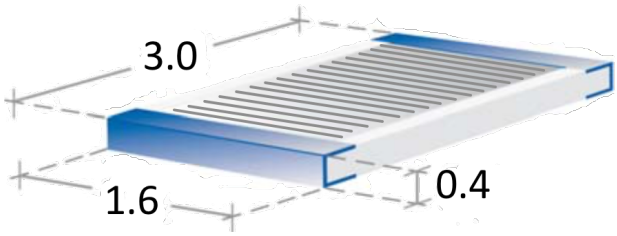


Figure 8.6: Image of a Pt100 RTD in a 1206 SMD package, with measurements in mm. The sensitive platinum track is vapour deposited onto the surface of the insulating ceramic substrate and protected by a few microns of inert glass. The conductive end caps solder directly to a PCB to connect into the measurement circuit

[1]Innovative Sensor Technology IST AG, Ebnat-Kappel, Switzerland
[2]IST GmBH Part Number P0K1.1206.2P.A, Farnell UK Part Number 1266933

Summary of Advantages

Comparing the characteristics of commercially available platinum RTD against those of thermocouples, as in Table 8.4 below, illustrates their greater suitability for the SmartPipe concept for monitoring biological liquids.

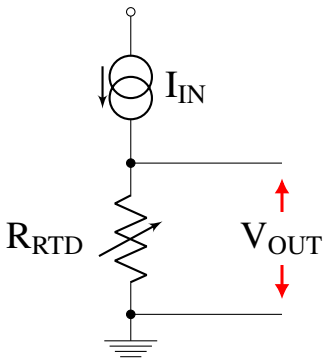
	Thermocouple	RTD
Advantages	Inexpensive Wide temperature range	Stability Linearity Accuracy Miniaturisability
Disadvantages	Non-linear Low sensitivity Low accuracy Zero-point compensation	Higher cost Self-heating Circuit complexity

Table 8.4: Summary of advantages and disadvantages of thermocouples vs. RTDs

8.3 Circuit Design for an RTD–Based Temperature Measurement System

Measurement Methodology

An RTD-based temperature measurement system relies on determining the absolute magnitude of the resistance with high precision.



$$V_{OUT} = I_{IN} \cdot R_{RTD}$$

In this simple circuit the voltage drop relates to the temperature (R_T) according to the

calibrated resistance of the device at 0°C ($R_\emptyset$) and the characteristic response curve of the RTD material according to its α and δ values.

H.L. Callendar (1863 – 1930) began his work on high-precision platinum resistance thermometers in the 1880s, and derived the following quadratic equation relating these characteristics from empirical measurements of α and δ at temperatures (T) ranging from 0°C to 260°C:

$$R_T = R_\emptyset \cdot (1 + AT + BT^2)$$

where:

$$A = \alpha + \frac{\alpha \cdot \delta}{100} \qquad B = \frac{-\alpha \cdot \delta}{100^2}$$

Therefore, deriving the temperature from the measured resistance of the RTD (R_T) is as straightforward as solving for T using the well-known quadratic formula:

$$T_{^\circ C} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Plugging in the known A and B coefficients:

$$= \frac{-A \pm \sqrt{A^2 + 4B \cdot (\frac{R_T}{R_\emptyset} - 1)}}{2B}$$

Table 8.5 below summarises the A and B values for the most common subtypes of Pt100 RTD:

α (ppm)	A	B
3926	3.9888×10^{-3}	-5.9150×10^{-7}
3911	3.9692×10^{-3}	-5.8495×10^{-7}
3850	3.9083×10^{-3}	-5.7750×10^{-7}
3750	3.8102×10^{-3}	-6.0187×10^{-7}

Table 8.5: A and B values for the most common Pt100 RTD materials, categorised by temperature coefficient, α. Note the values for the most common 3850ppm/K subtype highlighted in green

Measuring Temperatures Below Freezing

H.L. Callendar's equation was further modified by Milton S. Van Dusen (1892 – 1953) in 1925 to take into account a further empirical term β for nonlinearities at temperatures below 0°C, introducing a third-order 'C' term to Callendar's equation.

$$\mathrm{C} = \frac{-\alpha \cdot \beta}{100^4}$$

Resulting in the following form of the equation for temperatures below 0°C:

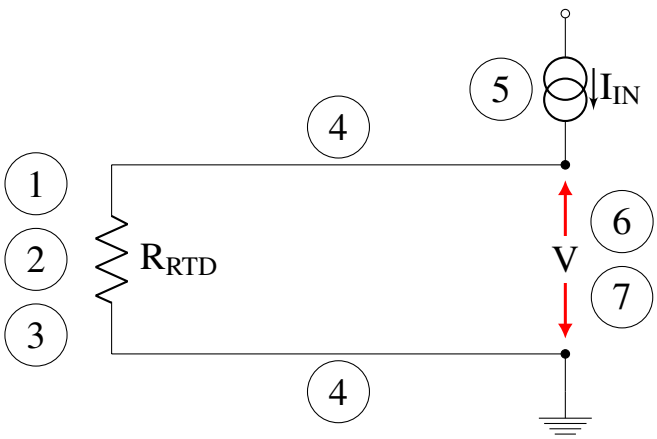
$$\mathrm{R_T} = \mathrm{R_\emptyset} \cdot (1 + \mathrm{AT} + \mathrm{BT}^2 + \mathrm{CT}^3)$$

This form of the equation will not be used because β, and therefore C, is 0 at temperatures at or above 0°C, and the perfusion circuit is not intended for use at temperatures below freezing.

Note that since 1968, the IEC has specified an even more complicated 20th order polynomial to fully characterise the temperature of an RTD across a full temperature range from 13K – 1200K. This is however an unnecessary range and degree of precision for the SmartPipe concept.

Specific Sources of Error in the Measurement Circuit

A high precision, high accuracy RTD sensor requires an equally well-designed measurement circuit, else the signal may be swamped by noise, drift or instability. This section considers sources of error which are specific to measuring the resistance of an RTD.



1. Systematic Errors
2. Thermal Gradients

3. Self-Heating
4. Line Resistance
5. Current Source Stability
6. Reference Stability

Systematic Errors

These result from slight impurities in the platinum alloy leading to deviations from the 3850ppm/K specification of IEC 60751 – hence the use of tolerance bands in the specification, given as the *Class* of the RTD sensor. The qualities of these Classes are shown here in Table 8.6, including that of the highly accurate[1] *Class Y*:

Class	Defined Tolerance ±°C	Deviation 0°C – 40°C
Y	0.10 + 0.0017×|T|	±0.168°C
A	0.15 + 0.002×|T|	±0.23°C
B	0.30 + 0.005×|T|	±0.5°C

Table 8.6: Maximum deviation of RTD measurement by sensor Class

The maximum deviation of an uncalibrated Class A sensor is just ±0.23°C across the working temperature range of 0°C – 40°C.

The second source of error arises the amount of platinum used to make the sensor. $R_{\emptyset}$, the end-to-end wire resistance of a Pt100 RTD at 0°C should be exactly 100Ω. However, due to manufacturing errors, this may not always be so.

In the case of the surface-mount chip RTDs that have been selected for the SmartPipe, their manufacturing process involves vapour deposition of platinum into tracks on the surface of a non-conductive ceramic substrate. The tracks are designed so that they may be laser-trimmed in post-processing to return the entire path length to as near to $R_{\emptyset}$ as is possible before protecting the surface with a few microns of inert glass.

Thermal Gradients

The active surface of a 1206 SMD RTD chip measures only 2.2mm x 1.6mm, and as such is unlikely to encounter any sufficiently steep thermal gradients in a flowing liquid. In the specification sheet of the surface mount RTD chips selected from IST AG, the chips reach 90% of their maximal response in flowing liquid within 450msec.

[1] And very hard to find

Self-Heating

Pulling a current across a resistor induces thermal (Joule) heating of the resistive material according to following equation:

$$\text{Power} = \text{I}^2 \cdot \text{R}$$

This thermal energy is usually dissipated into the environment, which may be either a flowing liquid or the underlying ceramic substrate in the case of the SmartPipe's chip RTD sensor. However, this thermal energy will also cause slight deviations in the total resistance of the RTD as a result of self-heating.

Assuming a current of 1mA flowing across a 'perfect' Class A 3850ppm/K Pt100 RTD at 20.0°C, the magnitude of the heating effect is as follows:

$$\begin{aligned} \text{R}_{20.0^\circ\text{C}} &= 107.7935\Omega \\ \text{Power} &= 0.001^2\text{Amperes} \times 107.7935\Omega \\ &\simeq 0.1\ \text{mW} \end{aligned}$$

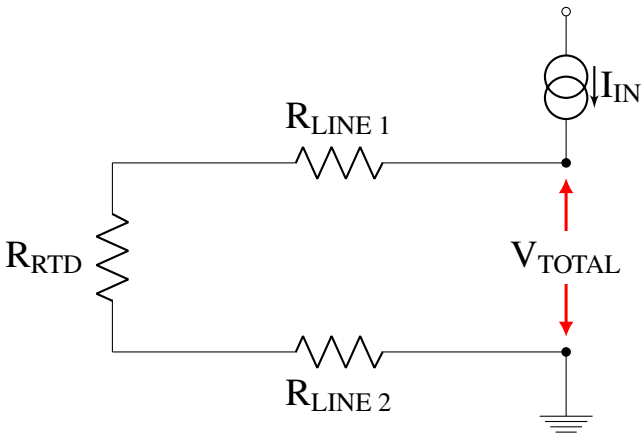
Giving approximately 0.1mW of thermal energy to be dissipated. According to the specification of the datasheet for the 1206 SMD RTD chip from IST AG, this will result in a thermal deviation of just 0.0018K in still water at 0°C.

Line Resistance

The RTD is not the only part of the circuit through which current flows and any line resistances will add an uncompensated offset to the total voltage drop, and as little as 40milliΩ can lead to an error of ≃0.1°C at 20°C (page 233).

2–Wire RTD Circuit

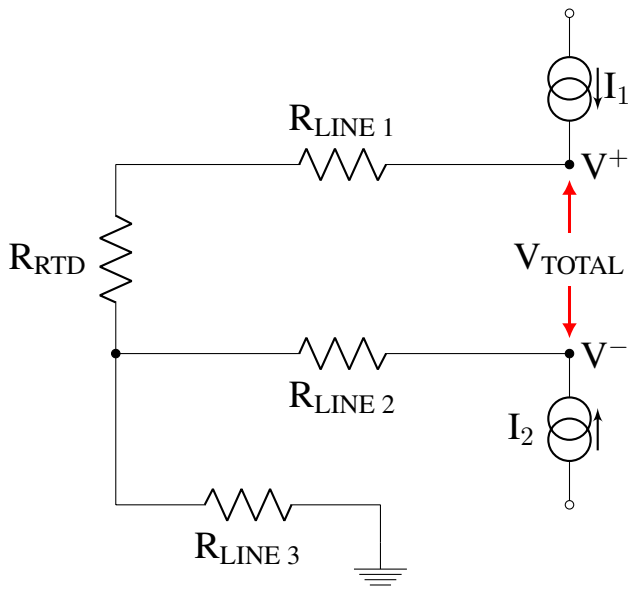
In the most basic circuit layout, the current source is attached in series with the RTD. This is simple to implement, but unfortunately the voltage drop includes the resistances of the outbound and return limbs of the circuit.



$$V_{TOTAL} = I_{IN} \cdot (R_{LINE\ 1} + R_{RTD} + R_{LINE\ 2})$$

3–Wire RTD Circuit

This implementation has one additional limb used to cancel out the average resistance of the other two leads, but it does require three perfectly matched lines and two perfectly matched current sources which can introduce additional errors and circuit complexity.



$$V^+ = (I_1 \cdot R_{LINE\ 1}) + (I_1 \cdot R_{RTD}) + (I_1 \cdot R_{LINE\ 3}) + (I_2 \cdot R_{LINE\ 3})$$
$$V^- = (I_1 \cdot R_{LINE\ 3}) + (I_2 \cdot R_{LINE\ 2}) + (I_2 \cdot R_{LINE\ 3})$$

If all three lines are equal in length and resistance:

$$V^+ = (I \cdot R_{LINE}) + (I \cdot R_{RTD}) + (I \cdot R_{LINE}) + (I \cdot R_{LINE})$$
$$V^- = (I \cdot R_{LINE}) + (I \cdot R_{LINE}) + (I \cdot R_{LINE})$$

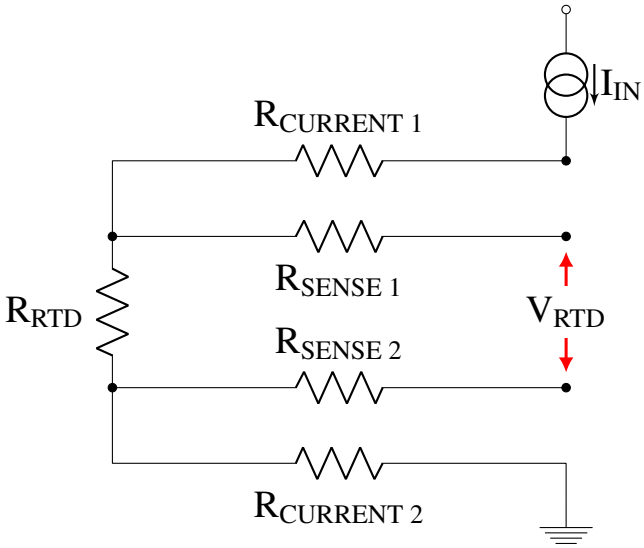
So:

$$V_{TOTAL} = V^{+} - V^{-}$$
$$V_{TOTAL} = (I \cdot R_{RTD})$$

4–Wire RTD Circuit (Kelvin Connection)

After considering the previous two circuit layouts, I decided on this particular design for a combination of precision and noise resistance.

One pair of traces forms a series loop passing current through the resistor, and the other parallel pair of traces reads the voltage drop across the RTD as a differential signal. This setup requires only one current source, unlike the two required for the 3-wire circuit.



With sufficiently high impedance at the inputs to the voltage sensor, there is no current flow to induce a voltage drop across their line resistances, and the measured resistance is solely a function of the current flowing in the circuit and the voltage dropped across the RTD:

$$V_{RTD} = I_{IN} \cdot R_{RTD}$$

Current Source Errors

An unstable current source can lead to errors in calculation. For example, measuring the resistance of a Class A 3850ppm/K Pt100 RTD with a ‘perfect’ 4-wire circuit (excluding the systematic error of $\pm 0.15°C + 0.002 \times |T_{°C}|$):

$$
\begin{aligned}
R_{20.0°C} &= 107.7935\Omega \\
R_{20.1°C} &= 107.8324\Omega \\
\Delta\Omega &= 0.0389\Omega \\
&= 0.499\% \text{ of } (R_{20.0°C} - R_{0.0°C})
\end{aligned}
$$

A 1mA excitation current must be in compliance by $\simeq 5\mu$A (0.499%) to ensure R_{RTD} doesn't approach the halfway transition point of 20.05°C between 20.00°C and 20.10°C. This error is compounded if two current sources are required for a precision measurement, hence the reason why I selected the 4–Wire RTD topology for the SmartPipe.

Fortunately, this potential source of error is abolished if the reference is derived from the same current source as the signal voltage, as is the case with *ratiometric referencing*.

Ratiometric Referencing

This is accomplished by generating the reference voltage from the same power source as the sensor signal, and allowing low frequency noise and power-line errors through. It is particularly well suited for resistive sensors requiring current excitation.

In the circuit shown in Figure 8.7 below, the digital code is derived from a ratio of the signal voltage across the sensor and that across a precision fixed-value resistor.

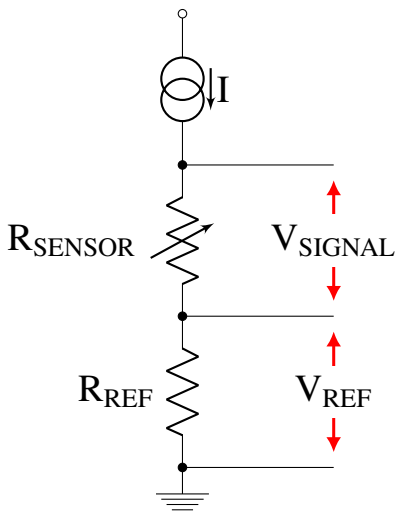


Figure 8.7: Layout of a ratiometric referencing circuit. The digital code is derived from a ratio of the signal voltage across the sensor and the voltage drop across a precision resistor of known fixed value

Here there are no semiconductor devices to produce shot noise or 1/f noise, and any changes

in the magnitude of the signal are *simultaneous* and of the same magnitude as shifts in the reference point. Noise, drift, and offset errors in the power supply are effectively ignored by this method. This ensures a linear relationship between the reference, the sensor voltage, and the digital conversion code.

The only remaining source of reference noise is the Johnson-Nyquist (thermal) noise, which for a 1kΩ reference resistor at room temperature with a signal bandwidth of 200Hz is on the order of just 57nV - acceptable even for a 24-bit ADC.

Note that this is also achieved at much lower cost and with far fewer components and board surface area than a comparable voltage reference IC.

The use of a ratiometric reference also greatly simplifies the calculations for the system designer. For example, the digital code produced from a unipolar 16-bit sensor is typically given as:

$$\mathrm{ADC_{OUT}} = \mathrm{GAIN} \cdot \left(\frac{\mathrm{V_{SENSOR}}}{\mathrm{V_{REF}}} \right) \cdot (2^{16} - 1)$$

$$= \mathrm{GAIN} \cdot \left(\frac{\mathrm{I_{SENSOR}} \cdot \mathrm{R_{SENSOR}}}{\mathrm{V_{REF}}} \right) \cdot (2^{16} - 1)$$

But if both $\mathrm{V_{SENSOR}}$ and $\mathrm{V_{REF}}$ derive from the same source, this simplifies to:

$$= \mathrm{GAIN} \cdot \left(\frac{\cancel{\mathrm{I_{SENSOR}}} \cdot \mathrm{R_{SENSOR}}}{\cancel{\mathrm{I_{SENSOR}}} \cdot \mathrm{R_{REF}}} \right) \cdot (2^{16} - 1)$$

Leaving the result based solely on the precision and stability of $\mathrm{R_{REF}}$, all else being equal:

$$= \mathrm{GAIN} \cdot \left(\frac{\mathrm{R_{SENSOR}}}{\mathrm{R_{REF}}} \right) \cdot (2^{16} - 1)$$

In this way the value of an LSB can be pre-calculated and hard-coded in software from the gain, bit-depth and $\mathrm{R_{REF}}$ values. This can accelerate calculations by the limited 8-bit ATMega328P microcontroller, or the implementation of look-up tables in memory - a subject covered in greater depth in Chapter 9 from page 317.

Selecting the Reference Resistor

To satisfy the requirements of a high accuracy, high stability voltage reference, the resistor itself must have (i) low temperature sensitivity, (ii) low drift, and (iii) high manufacturing tolerance.

Additionally, the value of the resistor must be as low as possible to minimise the Johnson-

Nyquist noise and the offset from any bias currents at the reference inputs, yet able to generate a sufficiently large voltage drop within the range of the ADC's reference inputs at the current specified for driving the sensor.

Fortunately, low cost precision surface mount resistors are available in a wide range of values with tolerances down to 0.01% and temperature coefficients $\leq$10ppm/°C.

Final Design Details

In summary, I have chosen to use an IEC60751 Pt100 RTD sensor in a 1206 surface-mount package for the SmartPipe, and a current-driven 4-wire circuit using the same current source to derive the reference voltage for the ADC in a ratiometric configuration. The voltage drop across the RTD will be measured as a differential signal and buffered to minimise any current drawn through the sensing lines.

The implementation and testing of the sensor will be covered in Chapter 9.

8.4 Selecting a Flow Sensor

The primary goal of the flow sensor in the SmartPipe is to aid the calculation of tissue oxygen delivery and consumption in the perfusion circuit, requiring a sensitivity of 10ml/min or greater over the range of 0 – 600ml/min.

I considered a number of diverse methods for measuring both volumetric and mass flow rate. Fortunately, the difference between these two is minimal in liquids with near unity density.

Turbines and Paddle Wheels

Figure 8.8: Roto-Flow™ paddle-wheel sensor. Bel-Art Products

These are the most common electronic flow sensors and rely on the liquid spinning a series of blades as it passes along a pipe. Because the volumes within the device are fixed and known, the flow rate can be derived from the rotational speed, measured non-invasively with Hall-effect sensors, or infrared reflectometry.

Whilst seemingly omnipresent in the catalogues of many commercial suppliers, this type of sensor did not appear to be easily available in the size, range or sensitivity demanded by the SmartPipe concept.

It initially appeared feasible to develop a rotation-based sensor of my own as the methodology was simple and eminently miniaturisable. Upon further consideration however, I found two main problems with the concept that rendered it unsuitable for the SmartPipe.

The most important of these was the requirement for physical blades within the flow path, risking hæmolysis and/or thromboembolisation with blood perfusion (see page 76 for a detailed discussion), and loss of pressure head.

Furthermore, the sensitivity of these devices are limited by the number of ‘counts’ per cycle, derived from the number of blades and the minimum liquid volume that will spin

them, which is then dependent on the quality of the bearings, the mass of the blade packet, and the form of the blades.

I therefore sought to investigate other, less-invasive methods of flow sensing.

Differential Pressure (Venturi Effect)

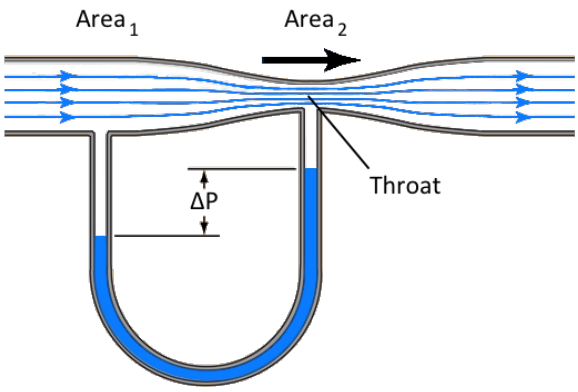


Figure 8.9: Illustration of the principle of the Venturi Effect flow sensor, where flow through a smaller diameter pipe must accelerate and exchange its pressure head for forward velocity

First described in 1797 by Giovanni Battista Venturi (1746 – 1822), the Venturi effect is the relative loss of pressure head as a fluid flows through a constriction point and pressure is exchanged for increased velocity.

With the cross-sectional areas of the normal pipe and the constriction point known, it is possible to derive the volume flow through the pipe by measuring this head loss. This can be done with two pressure sensors, or optical methods.

$$\text{Velocity} = \sqrt{\frac{2 \cdot \text{Pressure Drop}}{\rho \cdot (\text{Area Ratio}^2 - 1)}}$$

There is an additional correction factor applied to account for the 2%–10% head loss through the constriction that is dissipated as heat.

Whilst initially promising, this methodology unfortunately has many drawbacks that made it unsuitable for the SmartPipe concept.

The primary concern is the same as that for the more invasive paddle wheel or turbine sensors – a constriction in the 6mm diameter pipe sufficiently tight to provide a high area ratio and minimise the denominator would risk turbulence and hæmolysis (page 76).

Furthermore, the Venturi method is not sufficiently sensitive at the scale of the SmartPipe.

Table 8.7 below demonstrates the flow rates that correspond to just a 1mmHg pressure differential between the normal and constricted sections of pipe, a reasonable degree of precision for a sensor designed for biological pressures (page 161).

All values for density and pipe diameter have been taken from the values derived in Section 4.6.1, beginning on page 122.

Area Ratio	Radius (mm)	Volume Flow (ml/min)	Δ Pressure
2:1	2.121	14,580	1mmHg
5:1	1.342	5,500	1mmHg
10:1	0.948	2,700	1mmHg
100:1	0.300	269	1mmHg

Table 8.7: Flow rates corresponding to a pressure differential of 1mmHg, given a pipe inlet diameter of 6mm and fluid density of 1015kg/m^3. These results demonstrate the insensitivity of this method in the range required for the SmartPipe

These results demonstrate that even with a pipe constricting down to just 600μm in diameter, a Venturi flow sensor could not achieve the $\approx$10ml/min step size required to resolve a 1ml O_2/min difference in oxygen delivery.

Optical Particle Measurement

Optical measurements suffer from the requirements of high cost stable laser sources, fibre-optics, photomultipliers and associated electronics. They are not readily amenable to miniaturisation, nor incorporation into the SmartPipe.

Electromagnetic Methods

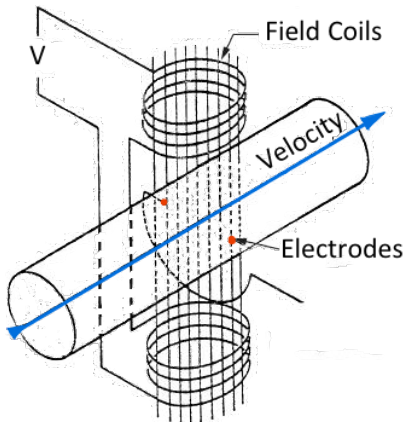


Figure 8.10: Electromagnetic flow sensor demonstrating the perpendicular arrangement of field coils and electrodes around the sample pipe

Because the perfusate will be a conductive, electrolyte-rich blood analogue, I briefly considered the possibility of designing an electromagnetic flow meter. These have no moving parts, do not interfere with the liquid under test, and create no pressure drop. Commercial EM flow meters claim full-scale accuracies up to 0.5%[1].

Electromagnetic flow sensors rely on Faraday's principle of induction, wherein the movement of conductors through a strong magnetic field generates a voltage in a receiver coil perpendicular to the static magnetic field. The voltage induced is proportional to the velocity of the electrolyte solution.

The electrodes must come into direct contact with the liquid under test, so as to measure the potential difference that develops between them. These could be miniaturised and manufactured from inert metals such as platinum or gold.

Unfortunately, after reading further into existing industrial designs it became apparent that I would not be able to use high-strength miniature neodymium rare-earth magnets as I had first imagined, because exposing the electrodes to a static magnetic field leads to polarisation and offset drift.

The magnetic field should instead be induced by AC field coils placed either side of the target pipe. Due to the dense packing of sensitive components with the SmartPipe, these high frequency switching coils would radiate noise into the rest of my sensors and risk degrading their performance, rendering this method unsuitable.

[1]Fluor Corporation, Irving, Texas, USA. Standards Certification Education & Training Presentation

Ultrasound

There are two main techniques for non-invasive ultrasound-based flow measurements. These are (i) time-of-flight, and (ii) doppler shift. Both of these methods are affected by alterations in the speed of sound in the liquid which can result from changes in temperature, density, and electrolyte concentration.

(i) Time-Of-Flight

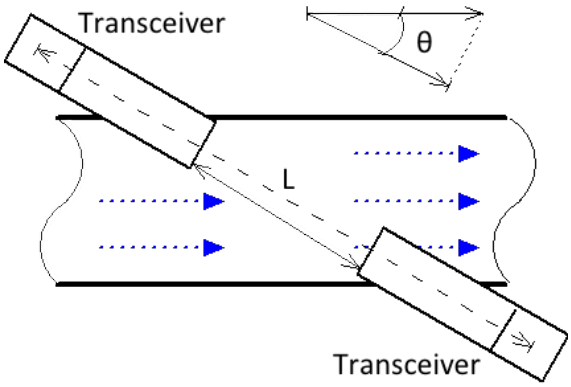


Figure 8.11: Time-of-flight flow sensor demonstrating the outbound ultrasound pulse, the angle between the sensor and the flow (θ) and the sensor separation (L)

This method relies on precise measurements of the time-of-flight of an ultrasonic beam between two piezoelectric elements. The downstream transducer emits pulses of ultrasound in the 1–2MHz range which are received by an upstream sensor placed at a known angle (θ) and distance (L) from the emitter. Alternately the emitter and receiver are co-planar and the signal is reflected from a plate on the far end of the device.

Because sound travels faster in the direction of a fluid flow than against it, a differential signal will develop in the presence of flow.

Using high-precision 'digital stopwatch' circuitry, the time-of-flight can be measured with nanosecond to microsecond precision, and fed into equation 8.1 to derive the mean velocity of the flow:

$$V_{\text{mean}} = \left(\frac{L}{2\cos(\theta)}\right) \cdot \left(\frac{t_{\text{up}} - t_{\text{down}}}{t_{\text{up}} \cdot t_{\text{down}}}\right) \tag{8.1}$$

This method *can* be used for particulate-containing flows, with the caveat that excessive scattering may attenuate the high frequency signal below the limit of detection. I therefore considered implementing the other form of ultrasound flow measurement.

(ii) Doppler Shift

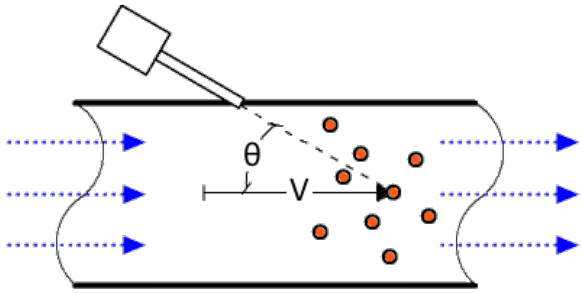


Figure 8.12: Doppler flow sensor demonstrating the signal reflecting off particles in motion (V) and the angle between the sensor and the flow (θ)

This methodology is better suited to flows containing particulates, such as my intended erythrocyte-containing perfusion medium or blood. Here, one piezoelectric element emits pulses at a fixed frequency (f_0), usually between 640kHz – 1MHz, and listens for the return signal.

This return signal will be frequency shifted (Δf) in direct proportion to the speed of sound through the liquid (c, ≈1575m/s in blood), the velocity of the particles (V), and the angle (θ) between the sensor and the flow, as described by equation 8.2 below:

$$V = \frac{\Delta f \cdot c}{2 \cdot f_0 \cos(\theta)} \tag{8.2}$$

The magnitude of the Doppler shift[1] is measured in Hz, and is positive if the object is approaching the emitter, and negative if receding.

Detecting changes as small as 1Hz on a carrier background of 1MHz should be within the realms of possibility for an RF mixer circuit, and driving the carrier frequency higher could further increase the sensitivity because blood scatters ultrasound proportional to the 4th power of frequency[129].

Each 1Hz Δf on a base frequency of 1MHz would be equivalent to the following change in blood flow rate with the sensor at 45° to the pipe, and :

[1]Named for Christian Andreas Doppler (1803 – 1853) for his work on describing the colours of observable binary stars in 1842

$$
\begin{aligned}
V &= \frac{\Delta f \cdot c}{2 \cdot f_0 \cos(\theta)} \\
&= \frac{1\text{Hz} \cdot 1575\text{m/s}}{2 \cdot 1,000,000\text{Hz} \cdot \cos(45)} \\
&= 0.0015 \text{ m/sec}
\end{aligned}
$$

Converting to volumetric flow through a 6mm diameter pipe:

$$
\begin{aligned}
Q &= 0.0015 \text{ m/sec} \times (0.003^2 \cdot \pi \text{ m}^2) \\
&= 4.24 \times 10^{-8} \text{ m}^3\text{/sec} \\
&\equiv 2.54 \text{ ml/min}
\end{aligned}
$$

However, I chose to abandon this method once it became clear that I would not be able to source sufficiently small piezoceramic elements or be able to implement the sensor in the SmartPipe in the time available for this Thesis. This was frustrating, as the method itself held great promise for a miniaturisable, high sensitivity, low-power flow sensor system with linear behaviour.

Instead, I decided to implement a miniature thermal mass flow sensor.

Thermal Mass Flow

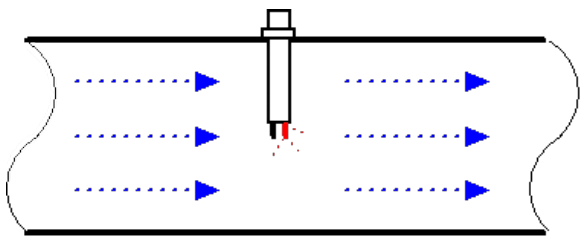


Figure 8.13: Illustration of the principle of the thermal mass flow sensor. Here the ambient temperature is sensed by the black probe, and the heater in red is driven to a specific over-temperature. The thermal energy required to maintain the over-temperature is proportional to the energy lost to the flow

Thermal mass flow sensors, or *constant temperature anemometers*, operate according to the principle of a continuous feedback loop attempting to maintain a heating element at a fixed temperature above that of a neighbouring ambient temperature sensor in the face of a fluid flow convecting this thermal energy away. The magnitude of the energy required to

maintain the temperature of the heater thereby corresponds to the mass flow rate of the liquid past the sensor and its heat capacity.

The reason for selecting this methodology was the immediate availability of the FS5L, a miniature combined solid-state sensor and actuator unit from the same manufacturer as the RTD chosen for the temperature sensor[1], at a cost of less than £10 per unit.

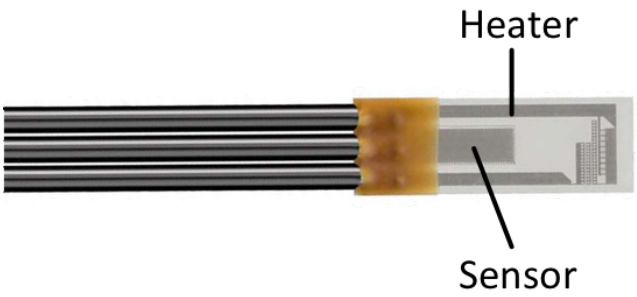


Figure 8.14: The FS5L solid-state thermal mass flow sensor from IST AG of Switzerland. The device incorporates both a peripheral heating element and central thermal sensor using platinum thin-film deposition technology

Figure 8.14 above does show a laser-trimming region to adjust the values of the elements plated onto the FS5L. However, the manufacturer has only specified the $R_{\emptyset}$ of the heater and sensor to $\pm 1\%$, which is far below the 0.12% of typical Class B RTD sensors.

Even with this shortcoming, the device is specified with a full-scale accuracy of $\leq 5\%$, a sensitive range of 1mm/sec – 10m/sec and operating temperature range exceeding the requirements of the SmartPipe.

Principle of Operation

Constant-temperature flow sensors are implemented as bridge circuits with the sensor and actuator occupying separate limbs, as shown in Figure 8.15 below, taken from the Application Note for the FS5L[130]. The documented $R_{\emptyset}$ values of the sensor and heater are 1200Ω and 45Ω, respectively.

[1]Innovative Sensor Technology IST AG, Ebnat-Kappel, Switzerland

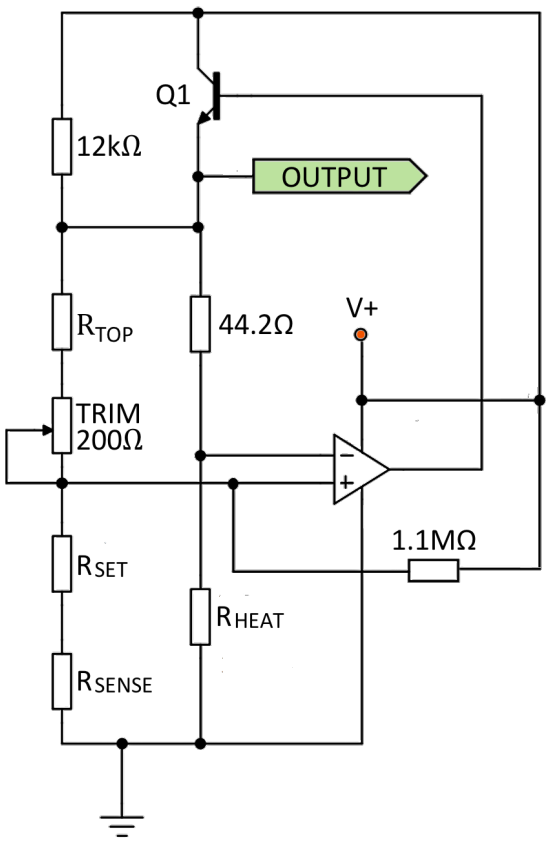


Figure 8.15: Bridge circuit for the FS5L thermal mass flow sensor showing the heater (R_{HEAT}), sensor (R_{SENSE}) and temperature offset resistor (R_{SET}). The value of R_{TOP} should match that of R_{SENSE}, and upper half of the limb is adjusted by the trimmer to match R_{SENSE}+R_{SET} at zero flow

In this arrangement, the higher resistance of the sensor limb draws a small current to create a voltage drop in the limb. This will be proportional to the resistance of the sensing element and hence the ambient temperature *plus* an offset introduced by the fixed-value R_{SET} resistor. This feeds into the non-inverting input of a difference amplifier as the set-point. The lower resistance of the actuator limb passes more current and heats the element by Joule heating.

The error voltage is sensed by the difference amplifier which controls the base current of the transistor at `Q1`, mainly increasing the current through the actuator limb until its voltage drop, and hence temperature, matches that of the sensor + fixed offset resistor.

The trimmer in the sensor limb is used to balance the circuit for the ambient temperature at zero flow.

The 1.1MΩ resistor provides a small bias into the non-inverting input of the amplifier to keep the transistor open, and the 12kΩ resistor across the transistor provides a small continuous leak current to activate the sensor.

Section 8.2 (page 222) discusses the principles of platinum thin-film RTDs in greater

detail.

Analysis of the Thermal Mass Flow Sensor Circuit

This circuit had a few problems that appeared after deeper analysis. These were related to the (i) non-linear behaviour of transistors, (ii) the over-temperature set-point inaccuracy, and (iii) inaccuracy over the working range.

(i) Transistor Non-Linearity

Transistors are known to behave in a non-linear fashion at currents below the saturation point. Operating in this region, small changes in the base current driven by the amplifier may result in much greater collector currents than expected. This has important consequences for the sensitivity of the circuit to small changes in the imbalance of the two limbs, and its stability because of the lack of any negative feedback in the design.

(ii) Over-Temperature Set-Point Inaccuracy

Higher over-temperatures lead to greater sensitivity, with thermal gas flow sensors typically operating at over-temperatures of 30°C or more.

However, the cellular and proteinaceous components within blood may be damaged by heating to just 3°C over the normal human core temperature. I wished to explore whether it would be possible to re-engineer the circuit shown in Figure 8.15 to operate with an over-temperature of just 1°C.

For this reason I wrote a Monte-Carlo simulation in Matlab R2012b to explore the optimal value of the fixed offset resistor (R_{SET}) that would be required to generate this 1°C over-temperature.

The simulation was designed to mimic a real-world setup of the sensor circuit, with a single adjustment to the set-point of the trimmer in the sensing limb at 0°C and zero flow. The R_{SET} values were taken from the catalogue of a commercial supplier[1] to match resistors available with 0.1% tolerances in 0402 size between 10Ω – 50Ω.

The fixed constraints of the model were:

- The amplifier/transistor combination behave in a perfectly linear manner
- The inaccuracies of the sensor and heater *co-vary* to the same % magnitude
- The sensors behave as perfect 3902ppm/K RTDs
- The sensor limb resistor has been perfectly trimmed to match that of the sensor
- A 44.2Ω resistor in the heater limb – the nearest real-world value 1% below $R_{\emptyset}$ of R_{HEAT}
- System voltage is a stable +3.3V

[1] Permier Farnell plc, Leeds, UK

The model followed the algorithm below for each value of R_{SET}:

1. Randomise the following resistor values:
   - The Sensor and Heater resistors (R_{SENSE} and R_{HEAT}) to $\pm 1\%$ at $R_{\emptyset}$
   - The resistor above R_{HEAT} to $\pm 0.1\%$
   - R_{SET} to $\pm 0.1\%$
2. Match the R_{TOP} resistor above R_{SENSE} to the value of R_{SENSE}
3. Now step through the following steps for 0°C – 40°C in steps of 1°C:
   (a) Calculate the value of R_{SENSE} at the ambient temperature
   (b) Calculate the voltage into the amplifier from the sensor limb
   (c) Calculate the value of R_{HEAT} to meet this voltage
   (d) Calculate the temperature of R_{HEAT} at this resistance
   (e) Record the difference in this value from the target temperature

This was repeated 1,000 times to study the range of deviations from the target over-temperature at each value of R_{SET}. These results are presented as Figure 8.16 below:

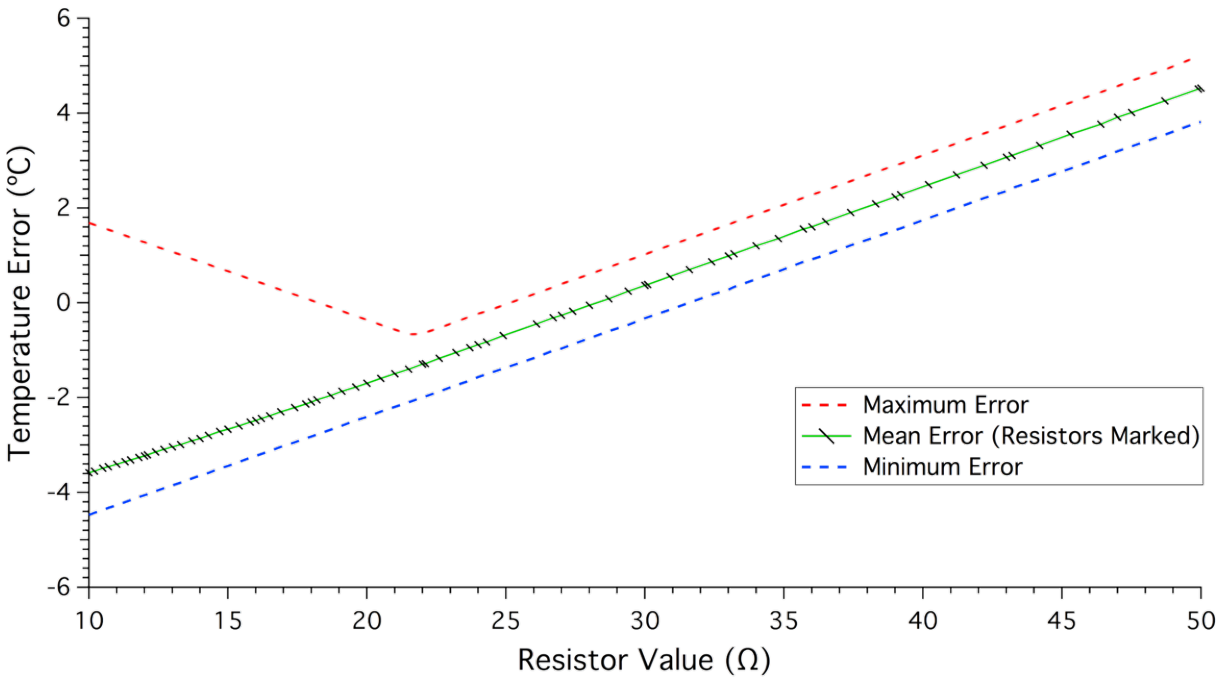


Figure 8.16: Results of the Monte-Carlo simulation of 1,000 circuits from page 244 with R_{SENSE} and R_{HEAT} covarying, and a perfectly trimmed resistor in the R_{SENSE} limb. Note the 'knee' effect from the occasional circuit where the upper part of the R_{SENSE} limb exceeds the value of R_{SENSE} + R_{SET}

This figure shows that all of the values fall within ±0.7°C of the mean, and that an R_{SET} of 28Ω will achieve the minimum deviation from the ideal 1°C over-temperature. Before this point, the maximum value shows the effects of the occasional randomly-generated circuit in which the upper part of the R_{SENSE} limb exceeds the value of R_{SENSE} + R_{SET} below the sensing point.

However, adjusting the parameters of the model to mimic the more likely real-world

behaviour of the system produces rather different results.

Figure 8.17 below demonstrates the effect of abandoning the trimmer, and instead using fixed value 1180Ω resistor accurate to 0.1% in the R_{SENSE} limb – the nearest real-world value 1% below the ideal $R_{\emptyset}$ of 1200Ω to accommodate the minimum value specified for R_{SENSE}.

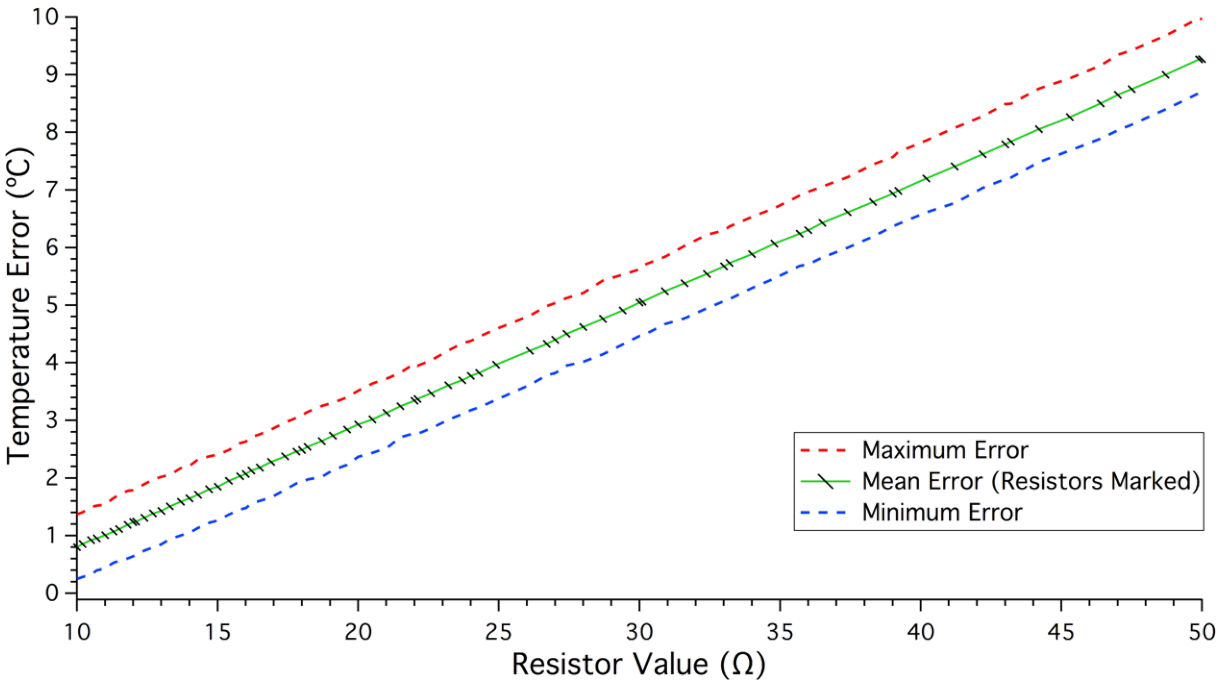


Figure 8.17: Results of the Monte-Carlo simulation of 1,000 circuits from page 244 with R_{SENSE} and R_{HEAT} covarying, and a fixed-value 1180Ω, 0.1% precision resistor in the R_{SENSE} limb

This tightens the error to within ±0.6°C of the mean, but even with the minimum available 10Ω resistor, the over-temperature would range from 1.25°C – 2.37°C, instead of the desired 1°C.

The reason for choosing a fixed resistor value just at the lowest $R_{\emptyset}$ value specified for R_{SENSE} can be demonstrated by further adjusting the model to allow R_{SENSE} and R_{HEAT} to vary independently within their 1% tolerance bands. Even with a 'perfect' trimmer in the R_{SENSE} limb, the accuracy of the system is now significantly degraded, as shown in Figure 8.18 below:

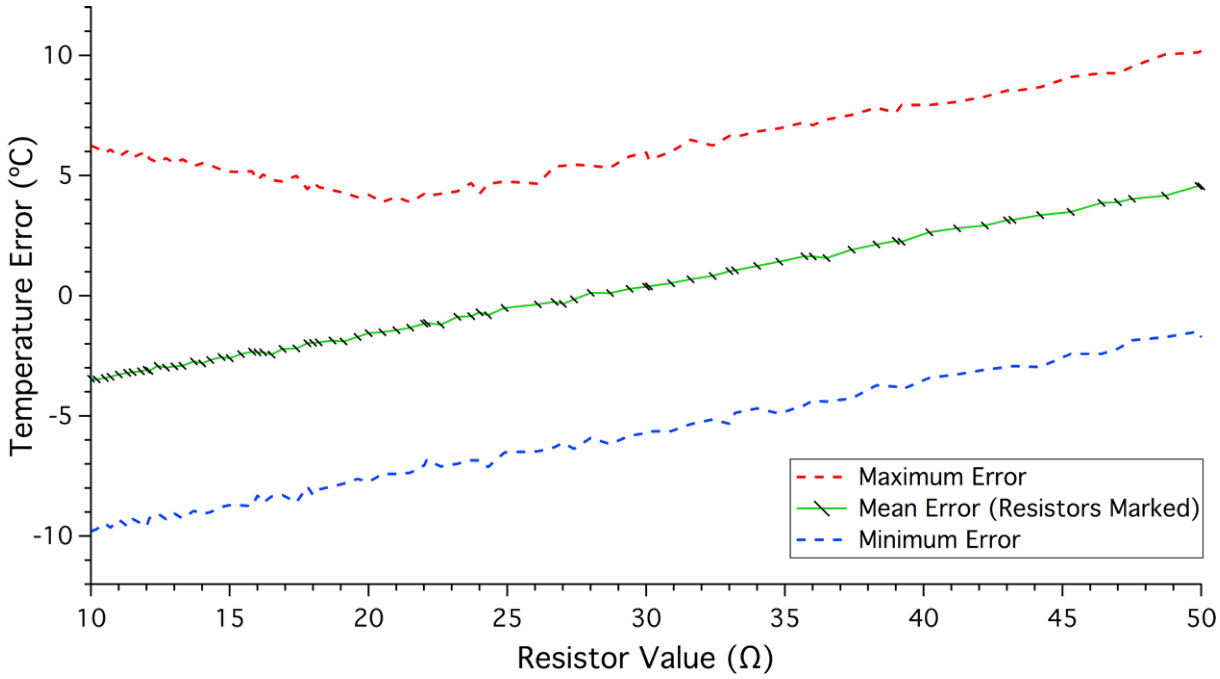


Figure 8.18: Results of the Monte-Carlo simulation of 1,000 circuits from page 244 with R_{SENSE} and R_{HEAT} varying independently within 1%, and a perfect trimmer in the R_{SENSE} limb

This figure is perhaps closer to the real-world behaviour of the system, as there is no guarantee from the manufacturer that the values of R_{SENSE} and R_{HEAT} co-vary. In this system, even at the optimal resistor value of 28.7Ω, the range of over-temperatures would range from 5.2°C *below* the ambient temperature, to 5.3°above the 1°C set-point.

The final figure in this series represents the similar results found by using the fixed-value 1180Ω resistor introduced previously. This is Figure 8.19 below:

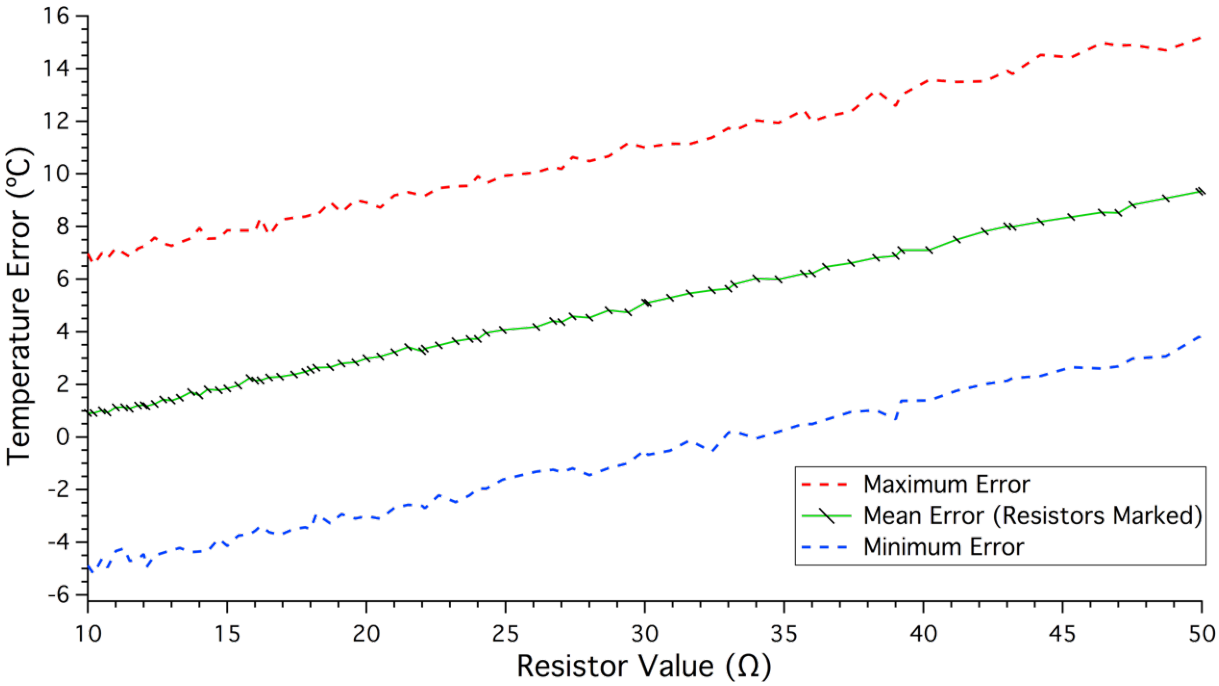


Figure 8.19: Results of the Monte-Carlo simulation of 1,000 circuits from page 244 with R_{SENSE} and R_{HEAT} varying independently, and a fixed-value 1180Ω, 0.1% precision resistor in the R_{SENSE} limb

This form of the circuit represents the likely range of errors in the over-temperature resulting from compounding the inaccuracies of real-world values of R_{SET} and all other resistors. The lowest error is seen with a resistor value of 10Ω, giving an over-temperature ranging from 3.9°C *below* the ambient temperature, to 7°C *above* the set-point.

(iii) Inaccuracy Over the Working Range

There is an additional problem with the analog circuit shown in Figure 8.15, which is the inaccuracy of the system over a working range following a single set-point adjustment.

So far, Figure 8.18 on page 248 has explored the inaccuracies of a circuit that has been trimmed perfectly at 0°C. Figure 8.19 on page 249 does not trim the circuit to balance the R_{SENSE} limb at 0°C but relies on a fixed resistor of 1180Ω.

Here we return to the simulation that produced Figure 8.18 and re-introduce a perfect trimmer, but trimmed at the working temperature of the circuit to explore the effect on the expected response across the operating range of 0°C – 40°C.

In the new Monte-Carlo simulation, R_{SENSE} and R_{HEAT} vary independently, R_{SET} is 28.7Ω with 0.1% accuracy, and the R_{SENSE} limb has been trimmed at 0°C. The simulation generated 1,000 virtual circuits, from which 10 were selected at random to produce Figure 8.21 below:

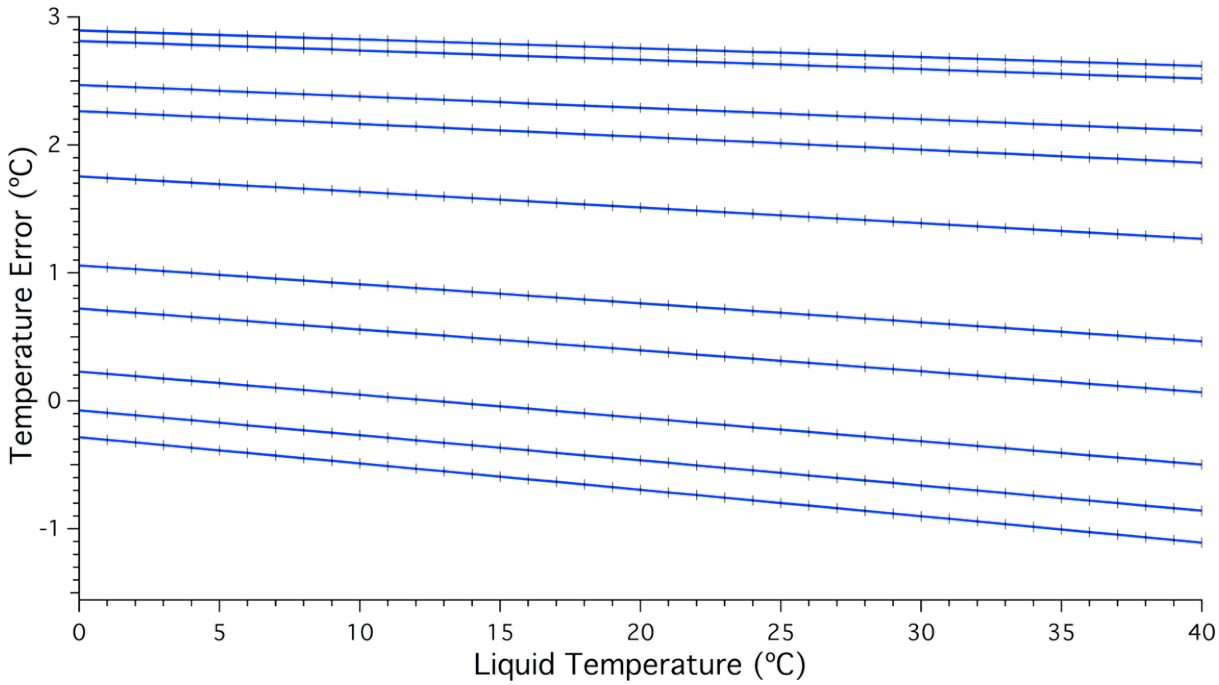


Figure 8.20: Deviation from the target over-temperature versus liquid temperature for 10 separate circuits produced by Monte-Carlo simulation of 1,000 circuits from page 244. The R_{SENSE} limb has been trimmed at 0°C

We see here that these 10 circuits selected at random behave within the bounds set by the simulation for Figure 8.18 on page 248, as expected. What was not initially apparent in the previous simulation was that the error magnitude *decreases* as the fluid temperature increases.

This same circuit with an R_{SET} of 28.7Ω was finally simulated as though it had been trimmed instead at 37°C. The results are shown below in Figure 8.21:

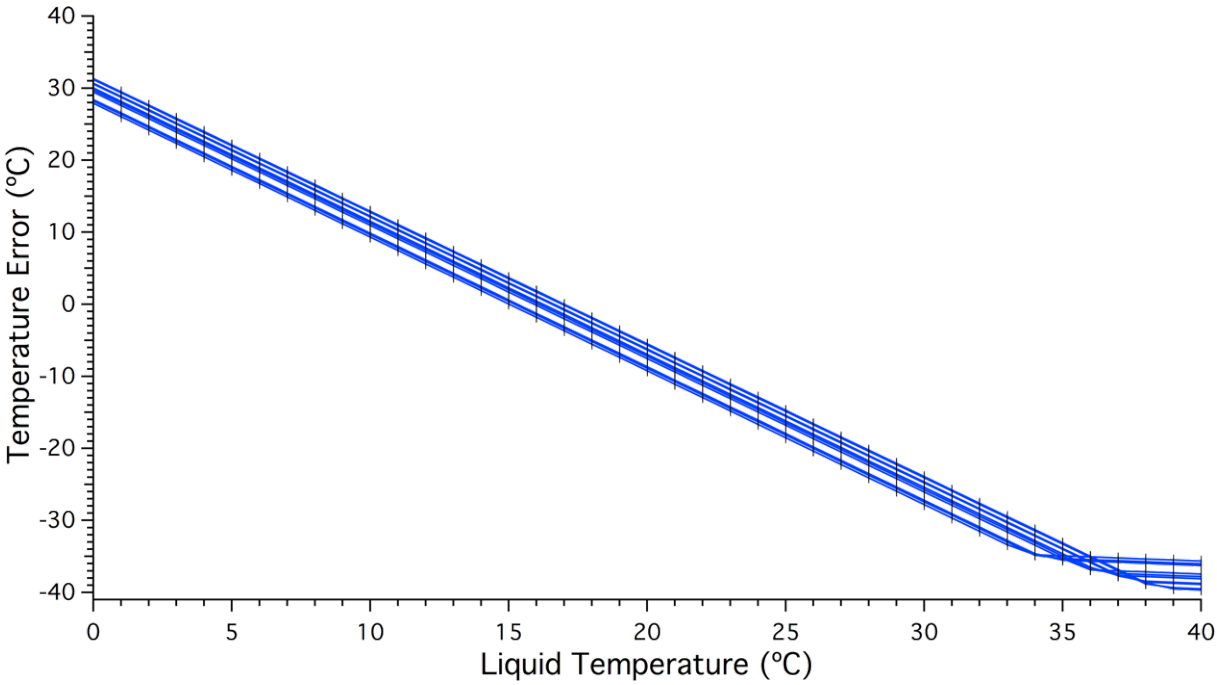


Figure 8.21: Deviation from the target over-temperature versus liquid temperature for 10 separate circuits produced by Monte-Carlo simulation of 1,000 circuits from page 244. The R_{SENSE} limb has been trimmed at 37°C. Note the 'knee' effect where the upper part of the R_{SENSE} limb exceeds the value of R_{SENSE} + R_{SET} until the trim point around 37°C

This simulation demonstrated that each working temperature requires a new adjustment of the trimmer in the R_{SENSE} limb, as well as the value of R_{SET}.

This would not be possible to implement in a straightforward manner to meet the goal of the SmartPipe operating as a largely turnkey system with minimal user input, nor could this circuit achieve the desired degree of accuracy and thus sensitivity whilst keeping the over-temperature at a safe level for blood flow.

Furthermore, after sourcing 5 of these devices direct from the manufacturer for testing, the liquid sensing version was withdrawn and re-classified solely for gas flow sensing owing to manufacturing difficulties with the insulating glass layer[1].

8.5 Circuit Design for a Digital Thermal Mass Flow Sensor

I decided to continue pursuing the thermal mass flow method, but to try to overcome these inaccuracies and the requirement for continually trimming the circuit and adjusting the

[1]Personal communication with the sales representative

set-point by using the precision of digital electronics and process control loops, with highly accurate RTD sensors in place of the FS5L.

Circuit Design

I first sought to combine a precision current sensor with an over-driven RTD working as a Joule heater, much like the 45Ω heater in the FS5L device. By measuring and controlling the current flowing through the circuit and the resultant voltage dropped across the RTD, I would be able to determine the resistance of the RTD and thus its temperature. A digital feedback loop would then regulate the current through the circuit in order to maintain this temperature at the desired over-temperature, with the current magnitude proportional to the flow rate.

In this way, the RTD would act as both the sensor and the actuator for the device, with the set-point temperature being fed in to the digital control system from a separate upstream RTD temperature sensor in the SmartPipe, or alternatively the sensor/actuator RTD could be switched between a high overdrive current and a lower sensing current to act as all three components – the ambient sensor, over-temperature sensor, and heater.

The first outline of this circuit is shown in Figure 8.22 below:

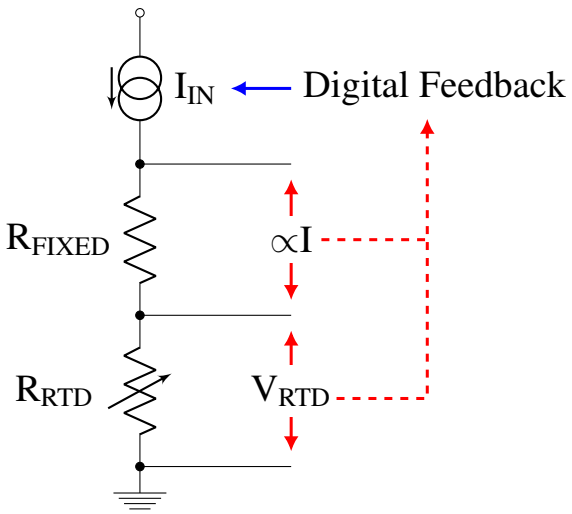


Figure 8.22: The initial design for the digital flow sensor. Both voltages are used to derive the resistance of the RTD and this value is then used to control the digital feedback loop

COMSOL Simulation

To explore the feasibility of this approach, I used a finite-element modelling package [1] to calculate the expected thermal energy required to maintain an over-temperature of 1°C with water flowing through a pipe in 10ml/min steps from 0ml/min – 500ml/min, matching the design of the SmartPipe. An image of the 3D model is shown in Figure 8.23 below:

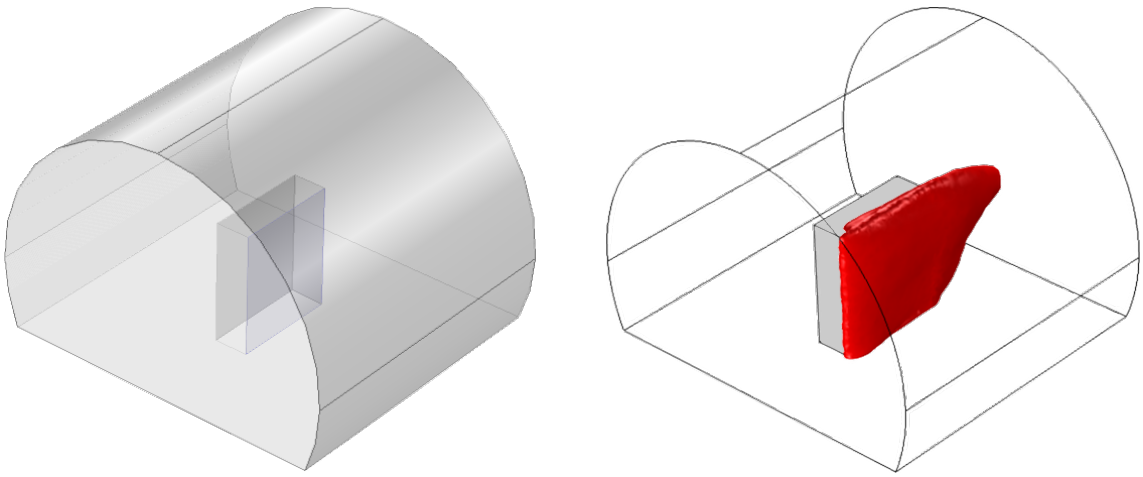


Figure 8.23: Left: Wireframe of the COMSOL model used to investigate the feasibility of an RTD-based, fully-digital thermal mass flow **Right:** Demonstration of the +1°C isosurface convected from the RTD at a flow rate of 500ml/min

The RTD was modelled as a commercially-available 50Ω RTD with a characteristic response (α) value of 3911ppm/K. The surrounding fluid within the pipe was chosen to be water, as the heat capacity and other temperature characteristics are similar to those of blood. The heating element of the RTD was presumed to be wholly within the flowing liquid, therefore preventing the heating of pipe itself which would lead to the signal drifting over time.

This study used a stationary non-isothermal flow model with a laminar inflow at the target flow rate, and the outflow at a mean arterial pressure of 90mmHg, the temperature of the ambient water was set to 0°C and 37°C to capture the potential working range of the sensor. One face of the RTD was assigned an over-temperature of 1°C, but the rest of the object did not have any thermal characteristics. The study was swept through 10ml/min incremental changes in the inlet flow rate across the normal working range of 0ml/min – 500ml/min.

The model was meshed within COMSOL without any additional refinements for edges or boundary layers. The program generated a free tetrahedral mesh with minimum element sizes of 17.3μm – 160μm, and 2 boundary layers around the RTD and internal aspects of the pipe. This produced a mesh with 194,000 elements at a minimum 'quality' of 0.2184,

[1] COMSOL Multiphysics Version 5.0, COMSOL AB, Stockholm, Sweden

exceeding COMSOL's own internal standards. Additional boundary layer refinements to 5 elements did not produce any improvements in the quality metric.

There were a few caveats to this model. An initial COMSOL simulation (not shown) used a 1206 Class A 100Ω RTD integrated into the wall of the pipe, as used for the temperature sensor (page 235). The simulations showed that this RTD value would have required supplies of 10V – 12V to produce the heating currents required, hence the reason for selecting a smaller magnitude 50Ω RTD.

Secondly, the requirement for the sensor to protrude into the flow stream went against my initial principles for the SmartPipe, but unfortunately there were no surface-mount RTDs with $R_{\emptyset}$ values of 50Ω. These could be custom ordered from manufacturers however, and so it seemed reasonable to proceed.

Finally, I did not include any thermal characteristics or coupling to the ceramic body of the RTD. Nevertheless, the wattage values derived from the simulation would only be increased with the additional heat loss through the ceramic substrate, thus making detection and control potentially easier.

From the model in Figure 8.23, I could derive the total normal heat flux from the surface in milliWatts. Using the known resistance of an RTD at 38°C I derived a close approximation of the current required to achieve this heat flux by Joule heating (I^2R).

The first results demonstrated that the voltage drop across the 50Ω RTD would reach a maximum value of 3.613V for an ambient temperature of 37°C and flow rate of 600ml/min, as shown in Figure 8.24 below:

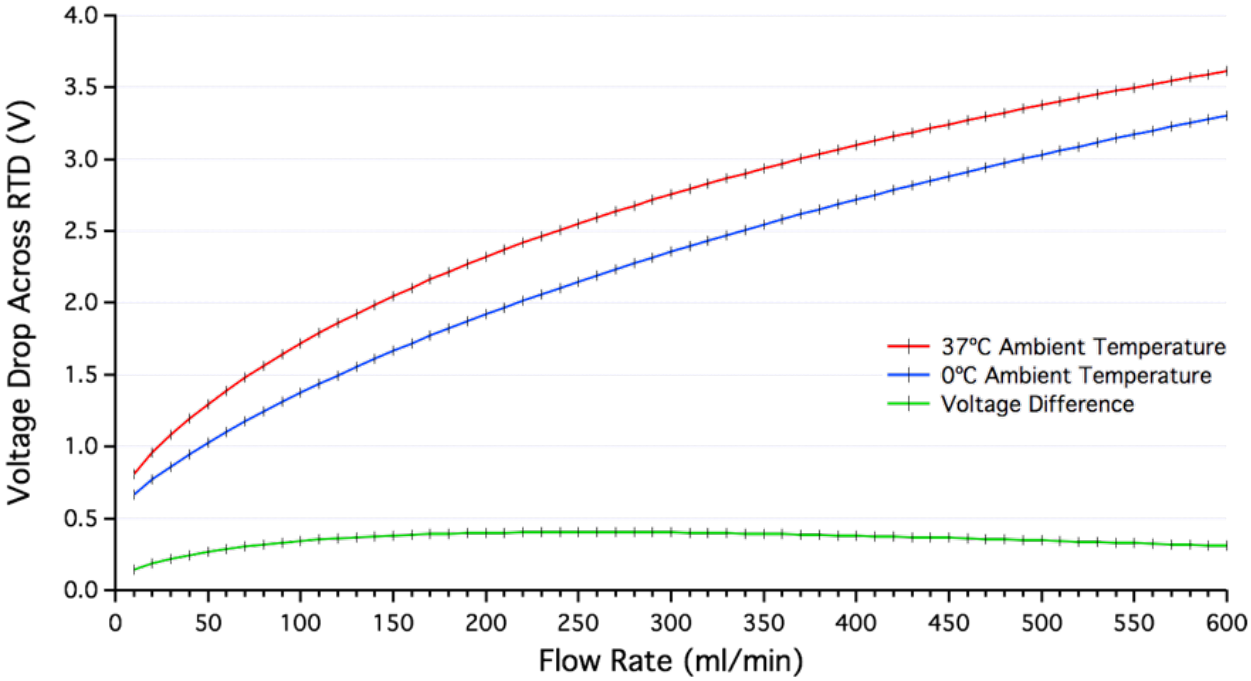


Figure 8.24: COMSOL model of the voltage drop across the 50Ω RTD to achieve a 1°C over-temperature for the given ambient temperature and flow rate, and the non-linear difference between these two curves

These two curves do appear to be separated by a fixed offset from approximately 150ml/min to 400ml/min when the difference begins to decrease again. This would indicate the need to create calibration curves for different background temperatures, or the derivation of an equation to solve the reverse problem of identifying the end-point of the control loop given the voltage drop across the RTD and the ambient temperature. However, and more importantly, they show that even the minimum 22mV step difference between 590ml/min – 600ml/min at 37°C lies within the capabilities of a 10-bit ADC.

The final result was to derive the necessary control currents to generate the fixed over-temperature at a given flow rate and ambient temperature. Figure 8.25 below demonstrates the current required to drive the RTD to 1°C over an ambient temperature of 37°C.

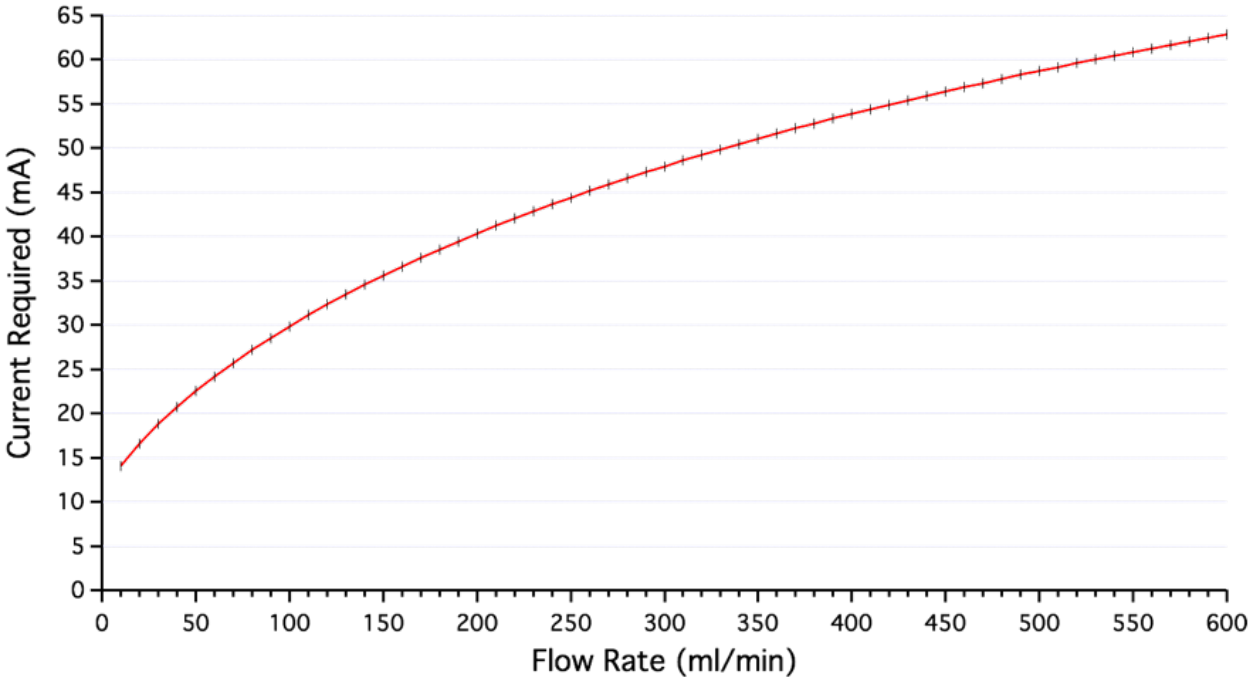


Figure 8.25: Graph of the current draw required to maintain a 50Ω RTD at 1°C above an ambient temperature of 37°C at different flow rates

The minimum current step here, again between 590ml/min and 600ml/min, is 380μA. This degree of digital current control was achieved with a combination of two components, discussed next.

Developing a Digitally-Controlled Current Source

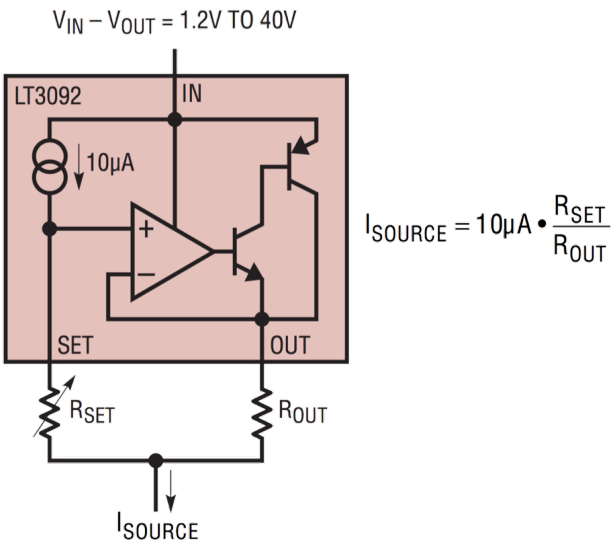


Figure 8.26: Diagram showing the internal design of the LT3092 showing a PNP/NPN transistor pair controlled by a difference amplifier, with the non-inverting input driven by the voltage drop generated across R_{SET} from a unique precision internal 10μA current source

This new approach required a miniature, controllable current source, for which I sourced the LT3092 device from Linear Technology[1] shown in Figure 8.26 above. This is a fully-integrated precision current source capable of delivering 0 – 200mA with 1% accuracy in a small 3mm×3mm DFN surface-mount package, requiring only two external resistors to control the current/voltage scaling.

I then required a miniature DAC to control the current through the LT3092, and settled on the AD5060 device from Analog Devices[2], a 16-bit high precision DAC with ±1LSB INL and buffered voltage output in a miniature 8-pin SOT23 package. The AD5060 was provided with a fixed 4.096V reference by a 16-bit accurate REF5040 device from Texas Instruments.

Combining the AD5060 and LT3092 led to the possibility of controlling the current source to the nearest 100μA. This required the LT3092 to be driven by the DAC in place of the LT3092's R_{SET} resistor, and using a 0.1% precision 10Ω R_{OUT} resistor to provide 100μA/10mV scaling.

Circuit Testing

To test the precision of the AD5060/LT3092 combination, 5 separate LT5092 devices in a mixture of 3mm×3mm DFN and larger SOT-223 surface mount packages were driven by AD5060 DACs controlled by an Arduino program I had written. Current and voltage data were collected using two Keithley[3] 2015 THD digital multimeters, one configured for recording the DC voltage from the AD5060, and the other to measure the current drawn by the LT3092.

The Arduino drove the LT3092 at 100μA steps between 0mA – 1mA, 1mA/step from 1mA – 10mA, then 10mA/step for the rest of the range to 100mA. There was an additional 100μA/step test at mid-range between 55mA – 55.5mA.

Figure 8.27 below shows the surprising linearity of the AD5060/LT3092 combination across the whole testing range:

[1] Linear Technology, Milpitas, California, USA
[2] Analog Devices Inc., Norwood, Massachusetts, USA
[3] Formerly Keithley Instruments, Cleveland, Ohio, USA, now part of Tektronix, Inc., Beaverton, Oregon, USA

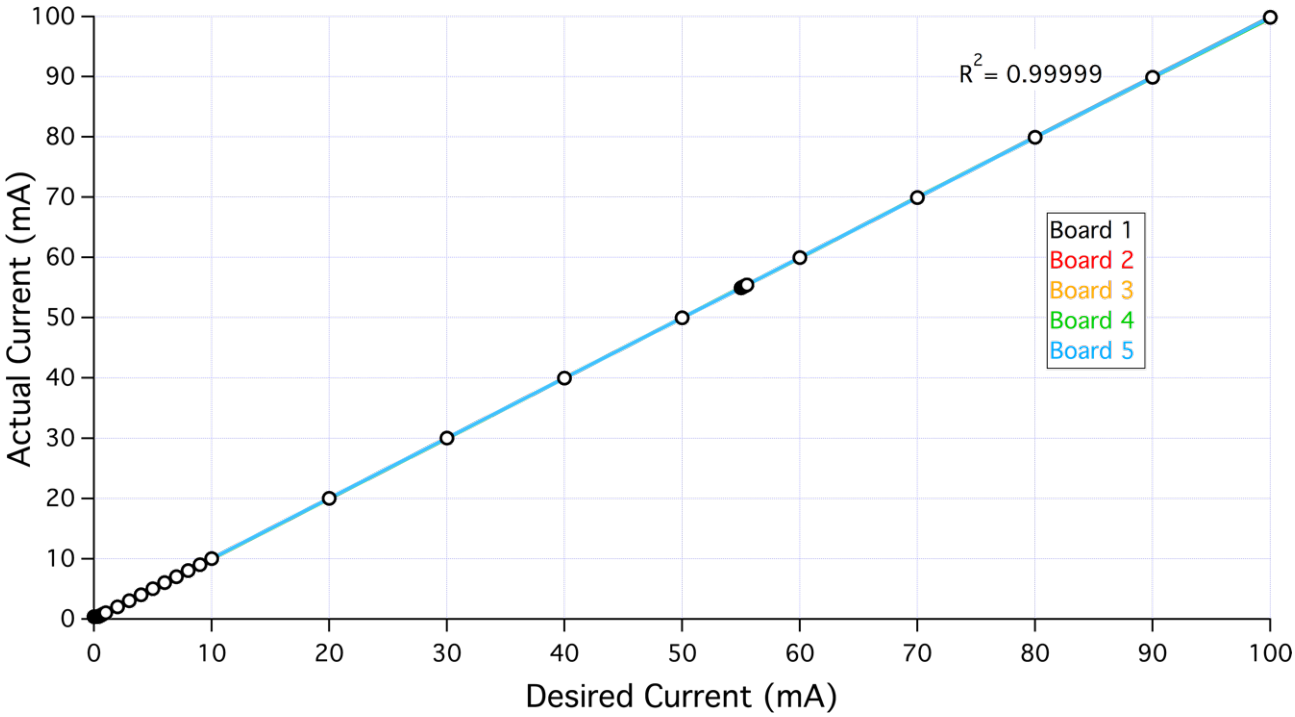


Figure 8.27: Testing the LT3092/AD5060 combination's ability to source currents between 0mA – 100mA

Zooming in on the 'knee' region around 1mA in Figure 8.28 below, we can see that this combination is unable to control a current below 400μA, and that each LT3092 appeared to have a fixed offset from ideal up to a maximum of 40μA.

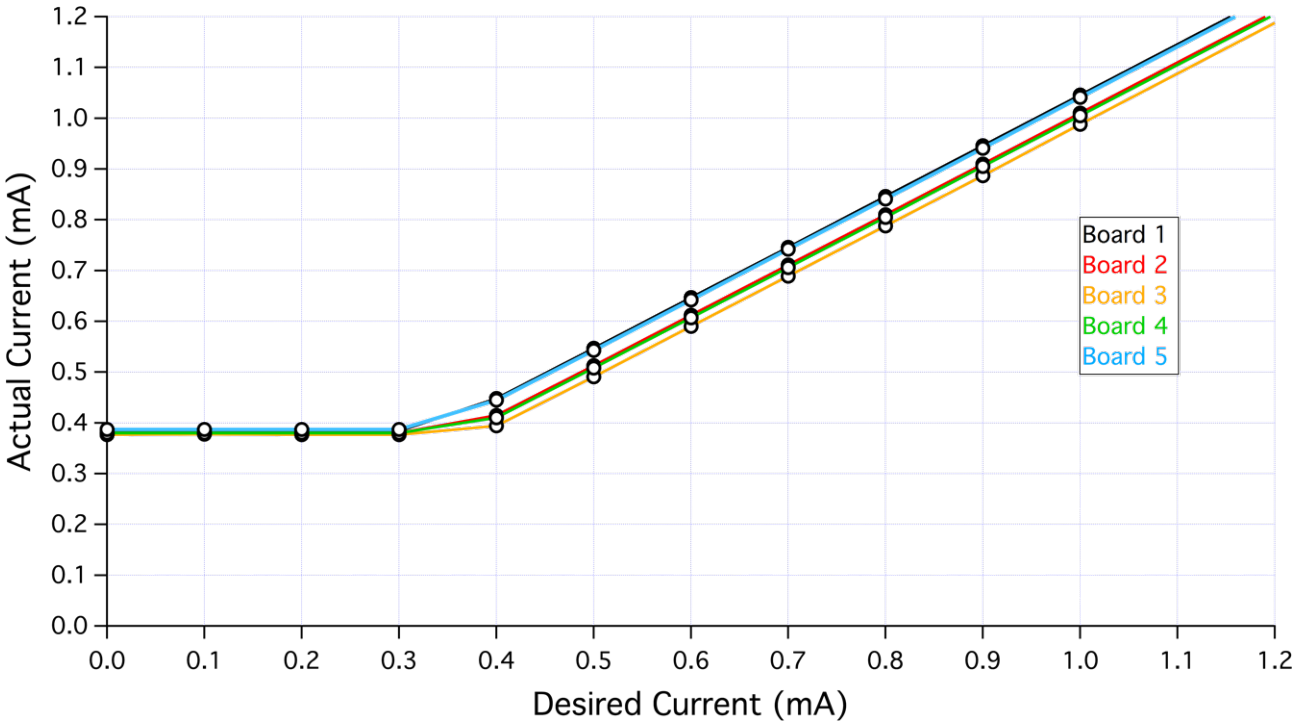


Figure 8.28: Close-up of the 0mA – 1mA section of Figure 8.27 demonstrating an inability to source currents below 400μA and fixed deviations up to a maximum of 40μA

Collating the data for the absolute error magnitude at each step demonstrated that the uncompensated LT3092/AD5060 system could be expected to source currents from 1mA – 80mA in steps of 100μA within an error band of 50μA, as shown in Figure 8.29 below:

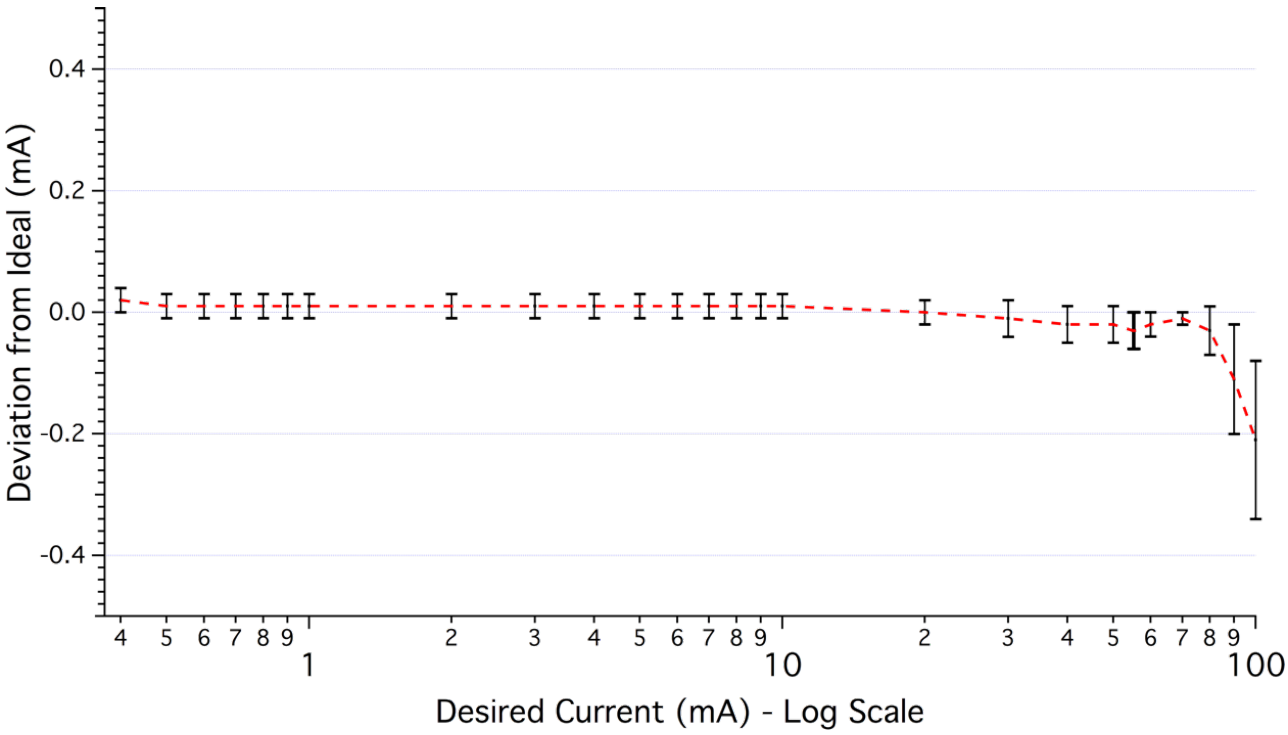


Figure 8.29: Absolute error over range for the uncompensated LT3092/AD5060 combination, demonstrating the deviation from the target current remains in a band of 50μA between 400μA – 80mA

Compensating this system with a resistor of known value in order to determine the fixed offset, for example by using the RTD's $R_{\emptyset}$ value at 0°C, could theoretically improve this tolerance further.

Second Circuit Design

With the accuracy of the LT3092/AD5060 suitable for use as both the sensor and actuator, the circuit shown in Figure 8.22 on page 252 was simplified to that shown in Figure 8.30 below, using an instrumentation amplifier (INA) to measure the voltage drop across the RTD:

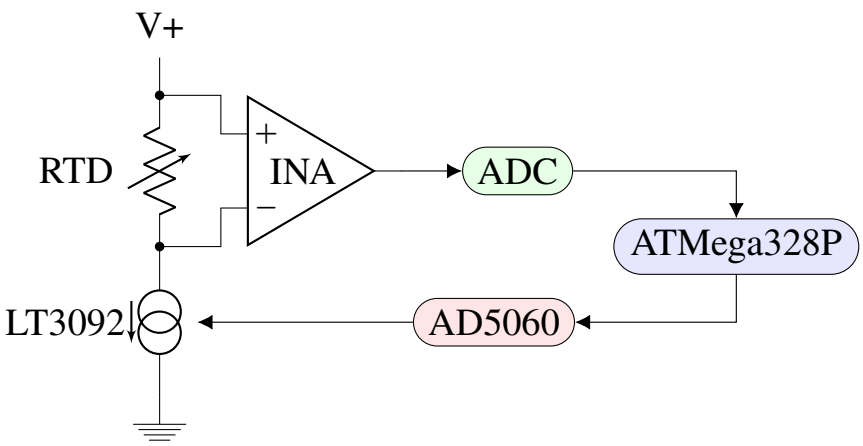


Figure 8.30: The final design for the digital flow sensor using the RTD as both the sensor and actuator, with a digital control loop to maintain the temperature of the RTD at 1°C above the ambient

In this circuit, the voltage drop across the RTD is sensed and digitised, feeding into the ATMega328P microcontroller which determines the correction signal required for the AD5060 to drive the LT3092 in order to hold the RTD's temperature 1°C above the ambient.

Concluding Remarks

These investigations have demonstrated that the standard analog methodology for creating a thermal mass flow sensor circuit by balancing two limbs of a bridge circuit is inadequately sensitive for low over-temperatures. This is largely due to the real-world tolerances of the various resistors required to balance the bridge, but the strongest influence comes from the tolerance of the resistances of the heating and sensing elements themselves.

I have proposed and investigated an alternative form of sensing circuit which relies on a high-precision RTD over-driven to act as a Joule heater, feeding into a digital control loop. The combination of a novel fully-integrated voltage-dependent current source (LT3092) and precision single-channel DAC (AD5060) is capable of driving the RTD heater with a high degree of accuracy, which could be further improved by a single-point offset correction against a known temperature.

Furthermore, this novel methodology for creating a digital thermal mass flow sensor allows the potential of implementing all three components – ambient sensor, feedback sensor, and actuator (heater) – as a single device. This could greatly reduce the size and cost of the system, whilst allowing for extremely low over-temperatures which can vary against a dynamic background, unlike the traditional system which requires modification of the set-point for each background temperature.

The greatest problem that I faced was the need for an instrumentation amplifier that was able to measure the voltage drop across from the supply rail across the RTD, as there are only a small range of commercially-available amplifiers with this capability. Attempts to produce an instrumentation amplifier from OpAmps able to sense up to their supply rails (e.g. the OPA192 from Texas Instruments) did not provide the required input impedance, but one innovation was to use a Schottky diode to provide a constant forward voltage drop from the supply rail to the RTD so that I could source an instrumentation amplifier that could instead sense to within 200mV of its supply. Unfortunately many of these devices have low output impedances and are unable to adequately drive a 16-bit accurate ADC.

My inability to complete the feedback loop most likely arose from my use of high-resolution, 16-bit ADCs for the sensing limb when selecting instead a lower precision, 10-bit ADC with integrated front-end components that better matched the requirements of the system would have been sufficient to detect the smallest required step-change in voltage and would have allowed me to complete the hardware sensing system.

8.6 Selecting a Pressure Sensor

As the molecules of a liquid or gas collide with each other or the walls of an enclosing container, there will be an equal and opposite reactive force exerted against the molecules in motion (Newton's third law). The frequency of collisions and the energy imparted with each collision directly relate to the force exerted on the surface with which the molecules are colliding. It is this property of a force applied per unit area which we term 'pressure'.

The ancient Romans developed the first concepts of pressure through their engineering of immense systems of aqueducts and lead piping to deliver water up to multi-storey *insulae*, creating siphoning mechanisms to traverse broad valleys and implementation of methods to reduce and regulate the pressure from 'mains' pipes to that required for domestic supplies, as evidenced in historic sites such as Pompeii[131].

It wasn't until the 1600s that figures such as Galileo Galilei (1564 - 1642), Evangelista Torricelli (1608 - 1647) and Blaise Pascal (1623 – 1662) began formally investigating the phenomenon of pressure once again.

The quantitation of pressure depends upon the selection of a zero reference point, leading to three different magnitudes that may be measured:

- Absolute pressure – referenced against the 'zero' of a theoretically perfect vacuum
- Gauge pressure – referenced against the 'zero' of local atmospheric pressure
- Differential pressure – referenced against another pressure measurement

Torricelli was the first to develop an apparatus that measured the *absolute* pressure, insofar as it was able to imagine a perfect vacuum at that time.

However, the more important pressure for biological systems is the *gauge* pressure, as human biological processes must necessarily take place at the bottom of our 'ocean of an atmosphere'[1].

This has led to an interesting historical hold-over in the medical field. Whereas scientific apparatus and worldwide laboratories have adopted the SI unit of the *Pascal* for quantifying pressure, the medical world remains steadfastly adherent to Torricelli's original units of measurement – millimetres of mercury, *mmHg*[2].

Hence the popular understanding that a 'good' blood pressure should be $^{120}/_{80}$, with implied units of mmHg.

[1] A phrase attributed to Torricelli
[2] 1 mmHg ≡ 133.322365 Pa at S.T.P.

Pressure Sensors

The principal measurement methodologies include (i) optical sensing, (ii) capacitative sensing, and (iii) resistive sensing.

Optical Pressure Sensors

There are three main methodologies used for optical pressure sensing[132] – (i) intensity-based measurements, (ii) strain measurements using Bragg diffraction gratings, and (iii) Fabry–Pérot interferometers.

Interferometric Fabry–Pérot sensors are the most commonly used optical pressure sensors in medicine due to their smaller size and temperature insensitivity. Here, a highly polished pair of glass or silicon plates are separated by a vacuum. An optical fibre carries laser light into the chamber through one of the plates. As the plate in contact with the medium is deformed by pressure in that medium, the return path of the laser light interferes with the incoming laser light, forming diffraction rings. Using an interferometer the magnitude of deflection and hence the pressure acting on the diaphragm can be determined.

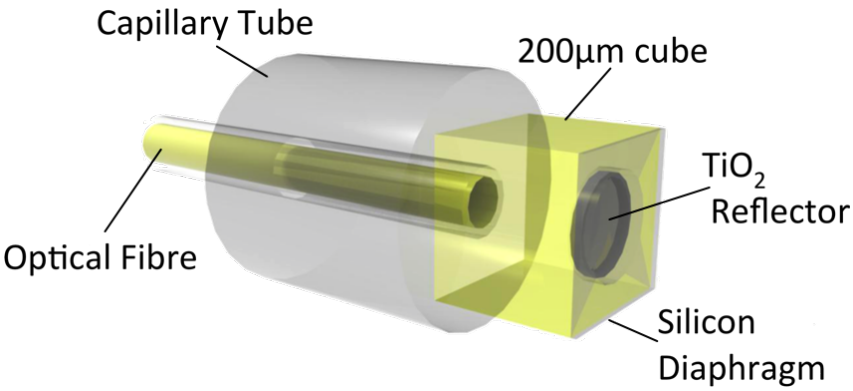


Figure 8.31: Fabry–Pérot interferometric pressure sensor design as proposed by Wolthuis *et al.* in 1991. From [132]

Whilst the probe tip itself can be very small in size (on the order of 125 μm in the case of the commercially-available F-125 sensor[1]), and suitable for use in liquids, they require costly and sizeable fibre optics and associated connectors as well as fast photodetection and post-processing circuitry in order to transduce the interferometric signal into a usable pressure tracing. Furthermore, although the sensitivity of the F-125 sensor is <1mmHg, its accuracy is at best ±5mmHg across the physiological pressure range of 0mmHg – 180mmHg.

[1]FISO Technologies Inc., Quebec City, Canada

It was therefore not possible to source a sufficiently low-cost, miniaturisable optical pressure detection element to integrate into the SmartPipe concept.

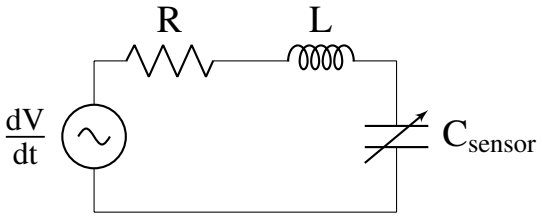
Capacitive Pressure Sensors

Capacitive pressure sensors rely on the formation of a variable capacitor with a fixed conductive plate and a second moveable element, with a dielectric of gas or gel between the two.

As pressure acts to deform the mobile element, the magnitude of the capacitance between the plates is altered, given by the formula for the capacitance between two parallel plates:

$$\mathrm{C} = \frac{\varepsilon_0 \varepsilon_r \cdot \mathrm{A}}{d} = \frac{\text{Dielectric Constant} \cdot \text{Plate Area}}{\text{Plate Separation}}$$

This can be detected by using the capacitive sensor as a variable capacitor in an AC-driven RLC circuit.



Here the frequency (ω) of the oscillating circuit reflects changes in the capacitance of the sensor according to the following relationship:

$$\omega = \frac{1}{\sqrt{\mathrm{L} \cdot \mathrm{C}_{\text{sensor}}}}$$

There is one capacitive pressure sensor that has been successfully trialled and licensed for clinical use in heart failure called the CardioMEMS™ HF System[1], but as a specialised medical device it is prohibitively expensive.

While the dynamic range of capacitive sensors can be very high owing to their mode of operation (often hundreds to thousands of mmHg), their lack of availability beyond prototypes demonstrated at conferences or costly specialised devices made them unsuitable for the SmartPipe concept.

[1] St. Jude Medical, Inc., Chelmsford, Massachusetts, USA

Resistive Pressure Sensors

Resistive pressure sensors are by far the most common type of pressure sensor in use today across all fields – industrial, automotive and medical – owing to the simplicity of their construction, robustness, and ease of integration with signal conditioning and detection circuitry.

Much like optical and capacitive pressure sensors, resistive pressure sensors monitor the deformation or movement of a diaphragm exposed to the medium being tested. However, resistive sensors do so by having conductive signal tracks as the sensing element. As the element is deformed, the end-to-end resistance of the track changes in proportion to the strain applied to the element, thus forming the simple *strain gauge*[1], as shown in Figure 8.32 below.

The principle underlying this change in resistance is the same as that underlying certain positive temperature coefficient (PTC) thermistors – the proximity of conductive elements within a matrix vary with its physical deformation.

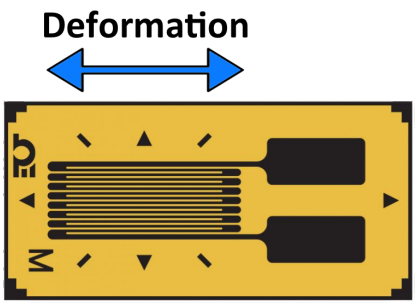


Figure 8.32: A typical 120Ω single-axis foil-bonded strain gauge which responds to longitudinal compression with a decrease in resistance, and to extension by an increase in resistance. Image from OMEGA® Engineering Inc., Stamford, Conneticut, USA

Resistive pressure sensors are then a form of strain gauge bonded to a deformable material such that the magnitude of deformation by pressure corresponds to the magnitude of the change in the resistance of the sensing elements.

Alongside the evolution of silicon microelectronics, techniques for the micro-manufacture for creating intricate *microelectromechanical systems* (MEMS) have been developed to fill the role of many previously macroscopic sensing elements. One leading use for MEMS has been in the miniaturisation of resistive pressure sensors, shrinking the sensing elements down to the μm scale.

[1] Strain is defined as the magnitude of the change in the shape or length of an element in response to a force

Piezoresistivity

MEMS pressure sensors rely on the phenomenon of *piezoresistivity*, the change in resistivity of a semiconductor crystal or metal to applied strain[1] to generate measurable responses in proportion to applied pressure.

Advantages of Using MEMS Piezoresistive Sensors

There are a number of reasons for choosing a resistive pressure sensor over other pressure sensing methodologies, including (i) sensitivity, (ii) precision, and (iii) ease of incorporation into the final design.

Sensitivity

With their greater degree of piezoresistivity, strain sensitive elements patterned from doped semiconductor materials can be up to 2 orders of magnitude more sensitive than bonded-foil gauges to the same degree of strain, allowing for miniaturisation of the sensors whilst still producing useable output signals. This effect is described by a relationship termed the *Gauge Factor* (γ), which describes the change in resistance per change in strain:

$$\gamma = \frac{\left(\dfrac{\Delta\text{Resistance}}{\text{Resistance}}\right)}{\left(\dfrac{\Delta\text{Length}}{\text{Length}}\right)}$$

Table 8.8 illustrates the far greater sensitivity of doped silicon compared to traditional materials used in strain-gauges.

Material	Gauge Factor (γ)
Manganin (Mn, Ni, Fe)	0.45 – 0.55
Constantan (Cu, Ni)	1.8 – 2.2
Silicon-based	110 – 200

Table 8.8: Gauge factors of common piezoresistive materials, demonstrating their relative sensitivities to deformation

[1] Unlike *piezoelectricity* which is a change in charge or voltage across a crystal to applied strain, and vice versa, the *piezoresistive* effect also occurs in non-semiconducting materials such as metals

Precision

With techniques borrowed from the world of integrated circuit manufacture, piezoresistive elements can be patterned onto a deformable silicon substrate to form a full MEMS pressure sensor from microscopic versions of the classic bonded-foil strain gauge shown in Figure 8.32. These MEMS devices also benefit from far tighter specifications and a greater degree of device-to-device reproducibility from these batch manufacturing techniques.

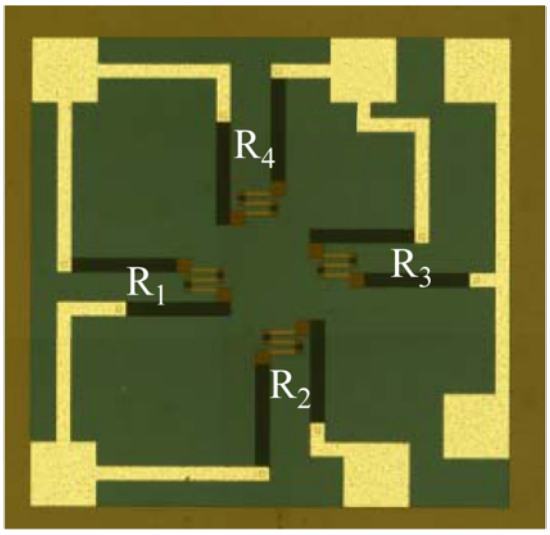


Figure 8.33: Detail of the die from a MEMS piezoresistive pressure sensor demonstrating four strain-sensitive elements (R_1–R_4) patterned onto a deformable substrate in resistive bridge circuit configuration. Image from [133]

Ease of Incorporation

Small MEMS piezoresistive pressure sensors are commercially available as medical-grade, sterile, disposable, fully-encapsulated units suitable for both liquid and gas-based pressure sensing, at around £6 each[1].

Two of the sensors considered were the NovaSensor® NPC-100 from Amphenol[2], and the 1620-4-N device from Measurement Specialties™[3], shown below in Figure 8.34. These are to all appearances identical products and so it was not possible to identify the original manufacturer. I selected the 1620-4-N on the pragmatic basis that it was the most easily sourced.

[1] For minimum orders of 20 units
[2] Amphenol Advanced Sensors, St. Mary's Pennsylvania, USA
[3] TE Connectivity Ltd., Schaffhausen, Switzerland

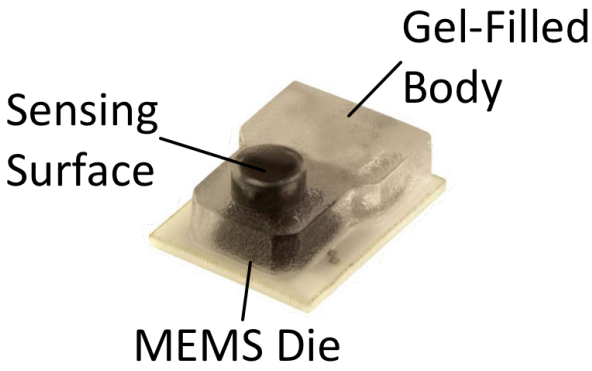


Figure 8.34: An image of the 1620-4-N MEMS-based pressure sensor from Measurement Specialties[3] demonstrating the 'chimney' capped by the cup-like sensing surface

These devices measure 10.54mm×8.13mm×2.26mm, with a protruding 1.27mm circular 'chimney' filled with an inert silicon gel, transducing pressure directly down to the MEMS die. This design means that the sterile, sealed 'chimney' can come into contact with the liquid to be monitored, while the body of the device remains outside in order to interface with signal conditioning and digitising circuitry which is far simpler than that required for optical pressure sensing.

The specifications for the device have been tailored for medical applications, with an operating range of -50mmHg to +300mmHg gauge pressure, a linearity and hysteresis across the operating range of ±1%, and an operating product life of 7 days. The manufacturers also claim their devices contain an automated temperature compensating mechanism.

One problem with the device was its unsuitability for reflow soldering, because the high temperatures required for reflow can melt the silicon gel filler and plastic body of the device. This necessitated a customised manufacturing process for the SmartPipe, detailed in Section 8.10 on page 283.

Summary of Advantages

In brief, the principal advantages to selecting a resistive pressure sensor over the other methodologies (optical and capacitive) are the availability of the devices at low cost, ready-tailored for medical applications across the physiological pressure and temperature range. Furthermore, the total size of the device and the complexity of the interface circuitry is greatly reduced.

8.7 Circuit Design for a Resistive Pressure Measurement System

Measurement Methodology

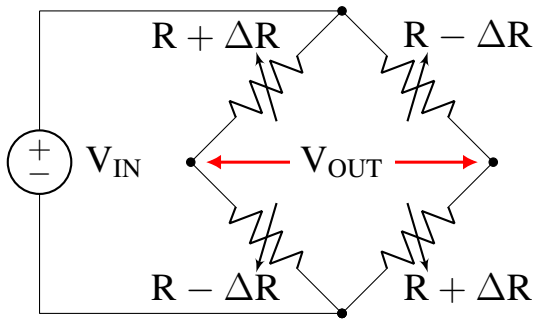
A piezoresistive pressure sensor records the *gauge* pressure of the system, and as such the absolute value of the resistance is less important than the magnitude of change from an established zero-point.

The fundamental methodology for measuring the resistance of the sensor is essentially the same as that for an RTD as described on page 226 – one passes a known current or voltage across a circuit and measures the other, with the ratio of proportionality being the *resistance*.

Four Element (Full) Bridge Circuit

The 1620-4-N sensor selected for the SmartPipe is internally implemented as a four-element, full-bridge circuit which overcomes many problems with measuring the values of single resistors in simpler divider circuits, including voltage offsets and non-linearity.

One example of such a full-bridge arrangement was shown previously in the figure of the MEMS pressure die on page 268.



With this design:

$$V_{OUT} = V_{IN} \cdot \left[\frac{\Delta R}{R}\right]$$

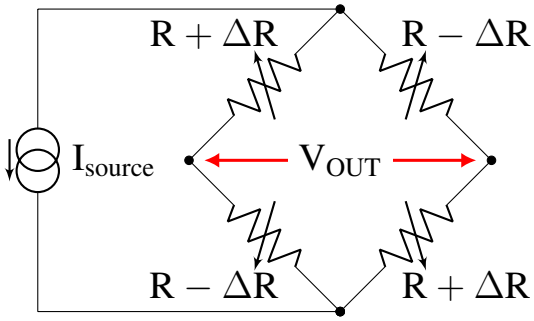
In this full-bridge circuit, the values of all four elements change with applied strain or pressure. The advantages of this configuration over all others includes:

- Twice the gain of a single variable element

- Linearity of the bridge output
- Zero offset
- Relative temperature insensitivity (all elements are affected equally by changes in temperature)

Current vs. Voltage Excitation

The linearity and noise performance of a full-bridge sensor can be further improved by driving the bridge with a constant current source, rather than a voltage source.



The output of the circuit is now simply:

$$V_{OUT} = I_{source} \cdot \Delta R$$

By using a constant current source to excite the bridge, and a differential amplifier with sufficiently high input impedance to measure V_{OUT}, no current will flow in the sense limbs of the circuit to induce errors due to line resistance or mismatch. Note that this also removes the role that the magnitude of the resistive elements play as the denominator, and only relies on the change in resistance with pressure.

With this method of driving the bridge, it is now also convenient to use the same current source to derive a ratiometric reference voltage for the system, as covered in Section 8.3 on page 233.

Specific Sources of Error

Manufacturing Tolerance

There are minor errors which can arise during the manufacturing process of the sensor, leading to mismatched resistances between the limbs or misalignment of the pressure transducing gel element such that there is a slightly unequal amount of force applied to

the four limbs of the sensor. The datasheet of the 1620-4-N device from Measurement Specialties™ describes the following tolerance characteristics:

Zero Offset:	±20mmHg
Offset Drift:	2mmHg over 8 hours maximum
Linearity:	±1% over working span
Hysteresis:	±1% over working span

Temperature Sensitivity

Despite the linearising effect of using a full bridge configuration driven by a constant current, there will be some imprecision introduced by thermal gradients, Joule heating and the *thermoelectric effect* (covered in more detail in Section 8.2 on page 218). Fortunately, the manufacturer has specified a thermal offset shift of just ± 0.3mmHg/°C, or a maximum of 11.1mmHg at 37°C which can be countered by ensuring that the zero offset is corrected at the desired operating temperature of the device.

Deriving the Value of the Pressure Sensor

Even with a current-mode full-bridge sensor, the absolute magnitude of the resistance is less important than the change from a set zero-point, and the use of a ratiometric referencing topology is for output stability and noise immunity.

Instead, the software for the pressure sensor uses a two-point calibration technique, deriving the zero-point from the ambient conditions and a secondary fixed point provided by the user and captured using the push-button interface (Chapter 7, page 204).

The system will then rely on the linearity and low hysteresis of the sensor and its circuit in order to derive a pressure reading.

Final Design Details

In summary, for the SmartPipe concept I have chosen to use the 1620-4-N medical-grade disposable pressure sensor and a current-driven circuit using the same source to derive the reference voltage for the ADC by ratiometric referencing, with a buffered differential amplifier conditioning the voltage signal from the bridge.

The implementation and testing of the sensor will be covered in Chapter 9.

8.8 Selecting a Blood Oxygenation Sensor

There are two main methods for determining blood oxygen content in clinical practice. These are (i) measuring the partial pressure of O_2 in the blood (pO_2), and (ii) measuring the amount of O_2 being carried bound to hæmoglobin (the O_2 *saturation* of the blood) with spectrophotometric methods.

Measuring the Partial Pressure of O_2

Recalling the form of the sigmoidal O_2-hæmoglobin binding curve from Figure 2.6 on page 64, and the design calculations from Section 5.5.3 on page 163, the amount of oxygen transported in solution only accounts for some ≈0.3% of the total at sea level.

Nevertheless, this O_2 is in equilibrium with the greater quantity bound to hæmoglobin within the erythrocytes. This means that if one is able to accurately determine the pO_2, it is possible to use a reverse look-up to determine the equivalent point on theO_2-hæmoglobin binding curve and thus derive the total O_2.

However, the true shape of the O_2-binding curve also shifts left or right under the influence of various factors including the pCO_2, pH, temperature and the presence of 2,3-bisphosphoglycerate (2,3-BPG), a byproduct of cellular metabolism. Determining the oxygen content of the blood by reversal of the pO_2 reading therefore requires multiple simultaneous measurements to correctly ascertain the form of the dissociation curve, including electrodes for the pH, pCO_2 and temperature sensing. This is particularly relevant to a normothermic organ undergoing perfusion, where the curve will likely be right-shifted by ischæmic metabolites to promote O_2 unloading.

I therefore considered adding the ability to sense pO_2 in addition to the more direct determination of the blood oxygen saturation by spectrophotometric methods.

Two techniques are in common use today for the measurement of pO_2 – (i) amperometry with Clark-type electrodes, and (ii) fluorescence quenching. Pulse oximetry is unable to derive the pO_2 from the sO_2 largely due to photon scattering effects.

Clark-Type Electrodes

Named after Leyland Clark (1918 – 2005) who successfully miniaturised and stabilised the electrochemical oxygen sensor in the 1950s for use in his blood oxygenation device, this methodology is now widely used in clinical blood gas analysers. Clark's electrode relies

on the direct reduction of oxygen at the surface of a suitably polarised platinum electrode[1] housed behind a protective membrane, as shown in Figure 8.35 below.

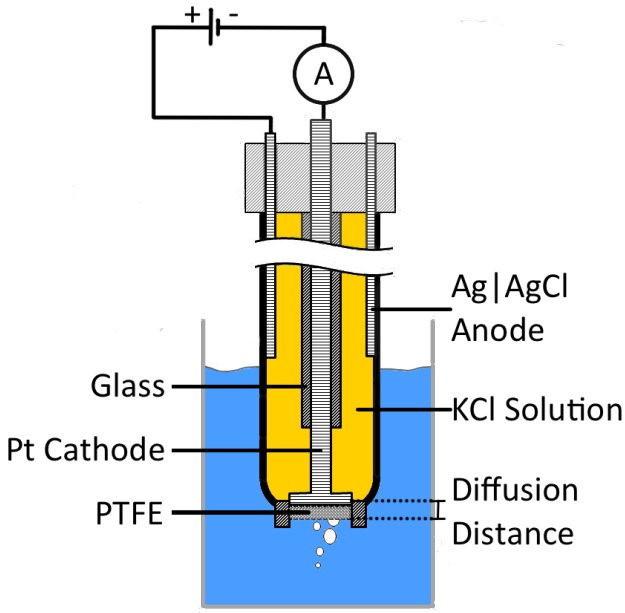


Figure 8.35: Image of a Clark-type electrode demonstrating the platinum cathode, Ag|AgCl anode in equilibrium with KCl, and the insulating PTFE (Teflon) membrane

In this arrangement, the electrode surface is protected from reducible species in the sample by a gas-permeable membrane. Ambient O_2 diffuses down the concentration gradient to the electrode surface, where it is reduced to water. The magnitude of the current (efflux of electrons) from the surface is then directly proportional to the pO_2.

$$\text{Ag Anode:} \quad 4\,\mathrm{Ag} + 4\,\mathrm{Cl}^- \longrightarrow 4\,\mathrm{AgCl} + 4\,\mathrm{e}^-$$
$$\text{Pt Cathode:} \quad \mathrm{O_2} + 4\,\mathrm{H}^+ + 4\,\mathrm{e}^- \longrightarrow 2\,\mathrm{H_2O}$$

There are of course problems with this method, including the need for a stable temperature, a suitably sensitive potentiostat, the potential for membrane fouling by blood constituents, and need to precisely control the membrane thickness to obtain a reasonable response time. It was this combination of potential problems along with the need to operate continuously for 24 hours or more within the SmartPipe, from where it could not be replaced, that led me to abandon this method of determining blood oxygen content.

[1] Usually -0.6V to -0.7V vs. Ag|AgCl

Fluorescence Quenching

This more recent technique relies on the ability of oxygen to reduce the fluorescence of complexes of certain platinum-group metals, most commonly ruthenium. One such 'reporter' molecule, $[Ru(dpp)_3]^{2+}$, is shown in Figure 8.36 below.

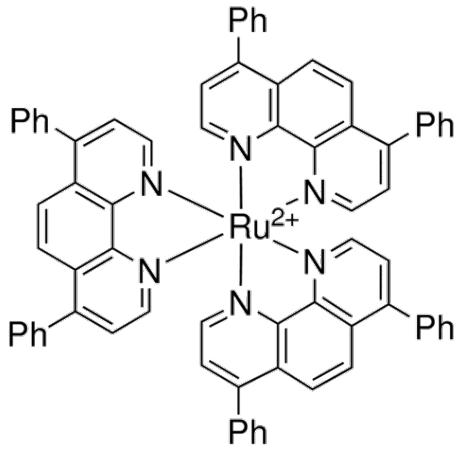


Figure 8.36: Image of tris(4,7-diphenyl-1,10-phenanthroline)-ruthenium(II), $[Ru(dpp)_3]^{2+}$, a common reporter for pO_2 measurement by fluorescence lifetime quenching

The ratio of fluorescence intensity in the presence and absence of O_2 (I and I_O, respectively) is proportional to the pO_2 and the Stern-Volmer quenching constant, K_{SV}, giving the Stern-Volmer equation:

$$\frac{I_O}{I} = 1 + (K_{SV} \cdot pO_2)$$

Thus, the sensing method requires the ability to detect the change in the fluorescence of a suitable 'reporter' species using a combination of fibre-optics, tuned laser sources at the excitation wavelength of the reporter (460nm for $[Ru(dpp)_3]^{2+}$) and fast photodetection circuitry tuned to the emission wavelength (613nm or 627nm for $[Ru(dpp)_3]^{2+}$). The CDI™ Blood Parameter Monitoring System 500 from Terumo[1] (page 156) employs this methodology in its fibre-optic remote sensor head.

Interestingly, $[Ru(dpp)_3]^{2+}$ has a relatively long fluorescence lifetime, on the order of μseconds. This has opened up the possibility of using less costly, lower precision, *non* fibre-optic techniques to detect quenching[134].

Rather than relying on precision measurements of the absolute value of the intensity, it is possible to use time-domain sensing to determine the phase lag between a regular

[1]Terumo Corporation, Shibuya, Tokyo, Japan

sinusoidal fluorescence excitation signal and the subsequent emission, which varies in proportion to pO_2[135]. This principle is shown in Figure 8.37 below.

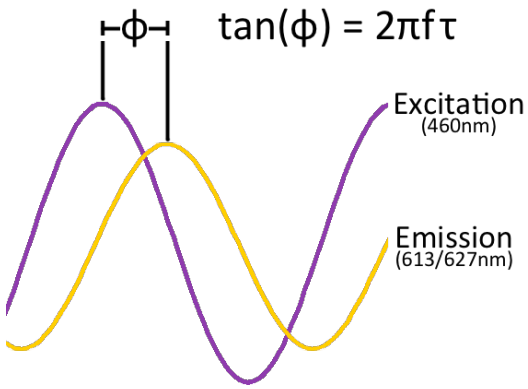


Figure 8.37: Diagram illustrating the principle behind phase fluorometry. ϕ represents the phase lag between excitation and emission, and τ the excitation lifetime of the fluorophore, which is independent of external conditions. The wavelengths given on the right are specific to the pO_2 sensitive $[Ru(dpp)_3]^{2+}$

Frequency-domain electronics operating in the low MHz range (corresponding to the μs lifetime of the $[Ru(dpp)_3]^{2+}$ reporter) are widely available and can be converted to measure phase differentials. There is even the AD8302, a single-chip, fully-integrated RF gain and phase detector available from Analog Devices [1], potentially allowing the entire system to be miniaturised and made solid-state.

However, I only discovered this particular method later in my project after having discounted the more complicated fluorescence intensity methodology. I decided at the time that it was too great a risk to abandon my progress with the spectrophotometric oxygen saturation sensor I was in the process of implementing.

Measuring the Oxygen Saturation of Blood

The second method of determining blood oxygen content is to directly measure the percentage of hæmoglobin binding sites that are occupied by O_2 through spectrophotometric differences between oxyhæmoglobin and deoxyhæmoglobin [2]. These non-invasive, non-consumptive methods have been in use for decades in clinical practice.

Figure 8.38 below shows the absorption spectra of oxygenated and deoxygenated hæmoglobin. Note that deoxyhæmoglobin absorbs more strongly in the red than

[1] Analog Devices, Inc., Norwood, Massachusetts, United States

[2] Arterial oxygen saturation determined by back-calculation from the pO_2 is termed the O_2sat, whereas the saturation derived from pulse oximetry is termed S_pO_2, and from spectophotometry (co-oximetry) of extracted blood the S_aO_2

oxyhæmoglobin, and this inverts in the near-infrared beyond the isosbestic point around 800nm.

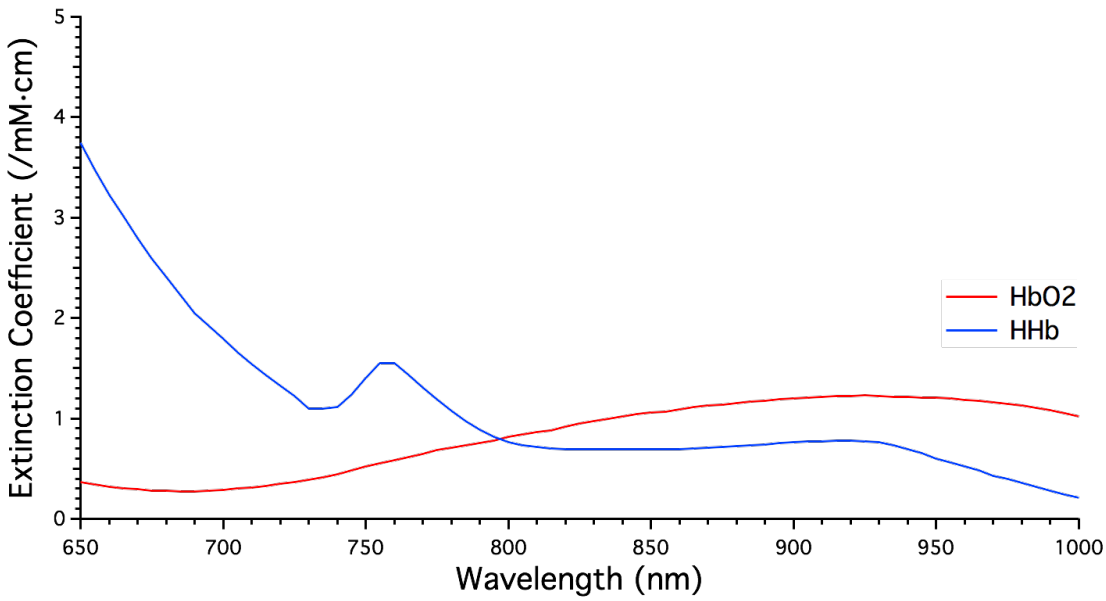


Figure 8.38: The absorption spectra of oxyhæmoglobin (red) and deoxyhæmoglobin (blue) between 650nm and 1040nm. Note the isosbestic point around 800nm. Data from [136]

This should allow the calculation of the sO_2 from the absorbance measurements at just two wavelengths of light (for example, 660nm and 905nm) and the known absorption coefficients (ε) by the application of the Beer-Lambert equation as follows:

Calculation of absorbances from intensity measurements:

$$\mathrm{A}_{660} = -\log \frac{I_{660}}{I_{o\ 660}}$$
$$\mathrm{A}_{905} = -\log \frac{I_{905}}{I_{o\ 905}}$$

Derive the photophore concentrations by solving the system:

$$\begin{bmatrix} \mathrm{HbO_2} \\ \mathrm{HHb} \end{bmatrix} = \begin{bmatrix} \mathrm{A}_{660} \\ \mathrm{A}_{905} \end{bmatrix} \begin{bmatrix} \varepsilon\mathrm{HbO_2}_{\ 660} & \varepsilon\mathrm{Hb}_{660} \\ \varepsilon\mathrm{HbO_2}_{\ 905} & \varepsilon\mathrm{Hb}_{905} \end{bmatrix}^{-1}$$

In practice, LEDs are not monochromatic, there are more than just two absorbing species in the blood, and neither the mean path length, scattering, nor the original intensity can be determined *a priori*. For this reason *pulse oximetry* is used clinically which accounts for these factors by examining changes in the red and infrared components against their static backgrounds over the cardiac cycle. In this way, the ratio-of-ratios internally normalises

the system, allowing for the use of calibration tables to derive the S_pO_2 to within 2% of the value determined by formal spectrophotometric analysis of extracted blood[137].

$$\text{Ratio} = \frac{\left(\dfrac{\text{AC}_{660\text{nm}}}{\text{DC}_{660\text{nm}}}\right)}{\left(\dfrac{\text{AC}_{905\text{nm}}}{\text{DC}_{905\text{nm}}}\right)}$$

And then applying the following linear approximation:

$$\%\ \text{S}_\text{p}\text{O}_2 = 110 - \left(\frac{\text{Ratio}}{4}\right)$$

Co-Oximetry

'Hæmoglobin' is actually a collection of multiple molecules with different O_2 affinities depending upon the species bound to the hæme moiety and the oxidation state of its coordinated iron atom. The metabolic byproducts hydrogen sulphide (H_2S) and carbon monoxide (CO) result in sulfhæmoglobin (SHb) and carboxyhæmoglobin (COHb) respectively. These are unable to bind O_2. Methæmoglobin (MetHb) results from the oxidation of the ferric iron in hæme to ferrous iron ($\text{Fe}^{2+} \longrightarrow \text{Fe}^{3+}$) preventing it from binding O_2 until reduced by a particular enzyme within the erythrocyte. The presence of the non-participatory hæmoglobins perhaps explain why the O_2 capacity of hæmoglobin has such a wide range (1.34O_2/g Hb – 1.39O_2/g Hb) in the literature.

The oxygen saturation is therefore more accurately described by the following equation:

$$\text{S}_\text{a}\text{O}_2 = \frac{\text{O}_2\text{Hb}}{\text{O}_2\text{Hb} + \text{HHb} + \text{COHb} + \text{SHb} + \text{MetHb}}$$

To derive the true O_2 capacity of the blood, the sensor system must be able determine the true proportion of oxygenated normal hæmoglobin vs. all other non-oxygenated hæmoglobins. This capability is termed *co-oximetry*. Figure 8.39 below demonstrates the absorption spectra of the major hæmoglobins – O_2Hb, HHb, COHb and SHb. High concentrations of MetHb are uncommon because of the endogenous enzymatic reduction back to HHb.

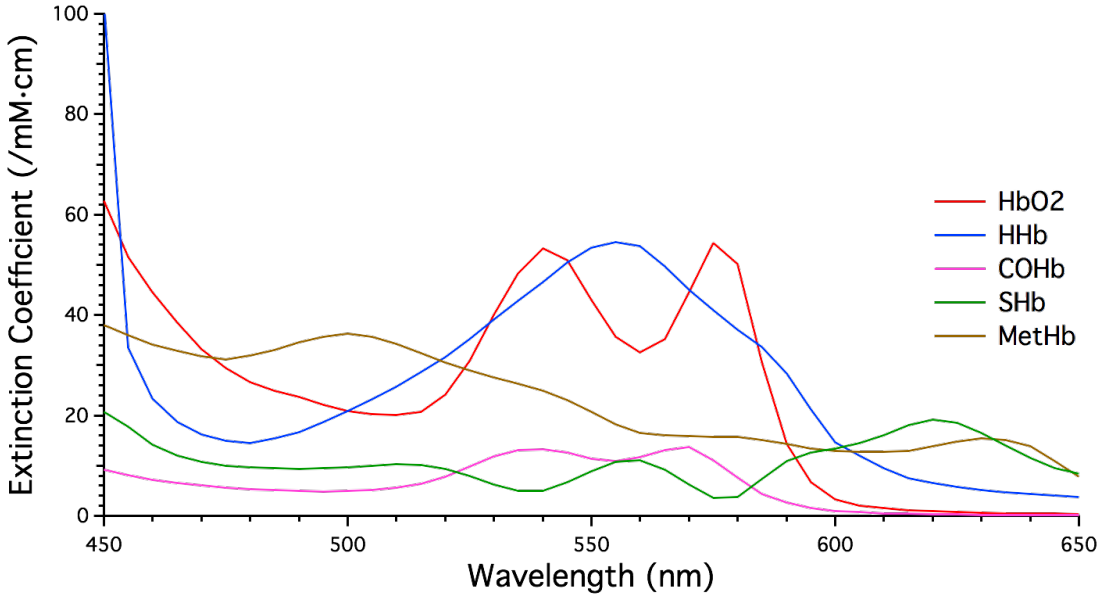


Figure 8.39: The absorption spectra of SHb, HHB, HbO2 and COHb between 450nm and 650nm. Data from the Natural Phenomena Simulation Group, University of Waterloo, Ontario, Canada

These four spectra would require a minimum of four wavelengths to probe the true composition of the blood. Commercial co-oximetry devices however can perform continuous broad-spectrum assessments of up to 128 wavelengths.

Designing the sO_2 Measurement System

During my initial research into potential designs I identified an OEM co-oximetry system which had the ability to determine all of the aforementioned hæmoglobin species, the total hæmoglobin content, and derive the total blood oxygen content. This was the MX-5 Rainbow® SET® from Masimo[1]. Unfortunately, this device required the purchase of a matching head-unit with no possibility for customisation or miniaturisation, and at a price in excess of £1,000 per unit, this did not meet the requirements for the SmartPipe.

Fortunately, we can simplify the requirements of the measurement system by assuming continuity between the arterial and venous limbs of the perfusion system and instead calculate the oxygen consumed by the target organ as a differential measurement between these limbs, as described in the design calculations on page 163 in Chapter 5.

Unless the kidney is significantly ischæmic, the hæmoglobin subtype profile should be consistent between the arterial and venous limbs, with the major difference being the proportion of $\frac{[O_2]}{[CO_2]}$. If the subtype profile does change however, this might produce

[1]Masimo Corporation, Irvine, California, USA

interesting research insights into the metabolic pathology of ischæmia and possibly new therapeutic targets.

As such I decided to create a simpler dual-wavelength oximeter for the initial version of the system. This would require me to source LEDs of the appropriate wavelengths (see Figure 8.38 on page 277) and a photodetector such as a photodiode, as well as designing and building a transimpedance amplifier stage and timing circuitry to correctly gate and match the sequential activation of each LED with the response of the photodetector.

Dual-Wavelength Probe

I identified and sourced a small, medical-grade, device containing two LEDs and matching photodiode from a small manufacturer based in Korea[1] at £10/unit. This front-end sensor has been specifically designed for measuring oxygen saturation and contains a 660nm and 905nm LED, each capable of 2mW output with a current of 20mA and a low forward voltage of 1.9V–2.2V. The device was packaged in a 9.8mm×4.3mm×1.3mm transparent plastic housing, and was suitable for surface mounting, matching the requirements of the SmartPipe.

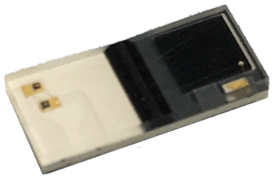


Figure 8.40: The surface-mount DCM03 sensor package containing one 660nm LED, one 905nm LED and a matched photodiode

The co-planarity of the emitter and detector means that the measurement relies on backscatter and reflection from the distal surface of the pipe, rather than the more typical method of measuring transmission between a LED and photodiode placed on opposite sides of the sample. However, reflection does increase the mean path length which in turn increases the signal attenuation and thus the sensitivity of the device.

Oxygen Saturation Circuit

Fortunately, at the same time as I identified the DCM03 in December 2012, Texas Instruments had recently released the AFE4490 device. This is a fully-integrated analog front end solution (AFE) containing a combination of LED drivers, timers, transimpedance

[1] APMKorea, Youseong Gu, Daejeon, Korea

amplifiers and analog conditioning circuitry and a 22-bit ADC. The device is available in a tiny 8mm×8mm 40-pin QFN surface-mount package for ≈£12/unit.

This device and the oxygen saturation sensor circuit will be discussed in greater detail in the next Chapter.

8.9 Final Design Details

Temperature
Methodology: Resistance Temperature Detector (RTD)
Sensor: IEC60751 Pt100 RTD sensor in a 1206 surface-mount package from IST AG
Circuit: Current-driven 4-wire circuit with ratiometric referencing

Pressure
Methodology: Piezoresistive MEMS bridge
Sensor: 1620-4-N medical-grade disposable sensor from Measurement Specialties
Circuit: Current-driven circuit with ratiometric referencing

Blood Oxygenation
Methodology: Dual-wavelength saturation measurement
Sensor: DCM03 dual-LED and photodiode sensor from APMKorea
Circuit: AFE4490 integrated analog-front end from Texas Instruments.

The implementation and testing of these sensors will be covered in Chapter 9

8.10 The Interface Pipe

The primary aim of the SmartPipe is to act as an low-cost multimodal sensing device that can fit unobtrusively into any perfusion circuit. One early decision to achieve this goal was to design the top board as a disposable unit, so that the sensitive and more costly electronic components on the middle and bottom layers could be preserved between uses.

Final Component Placement

Two major design constraints for integrating the sensors into the interface pipe were (i) the size of the DCM03 LED packet, and (ii) the inability to re-flow solder the pressure sensor.

The DCM03 measured 4.7mm in breadth and 1.3mm in height, which set the height required for the other components to lie flush. This required the use of two precision-cut 700μm tall copper stand-offs to raise the RTD to the level of the DCM03 after surface mounting.

These stand-offs were made to the size of individual 1206 pads and reflow soldered beneath the RTD at the same time as the DCM03. Their calculated resistance was $<$1milliΩ and would not therefore lead to significant offset in the RTD measurement circuit.

In order to accommodate the pressure sensor, I designed the PCB with a centrally-milled hole, and custom ordered from the manufacturer to a 2mm thickness. After epoxy mounting the sensor through this hole, the top of the 'chimney' came to sit at the same height as the DCM03 and the elevated RTD, so that the three sensors would sit flush with the inner diameter of the pipe.

Figure 8.41 below shows the top layer of the smartpipe with all components mounted, alongside an image of the underside of the PCB. The pressure sensor required hand-soldered wires to connect its pins to matching pads on the underside of the PCB.

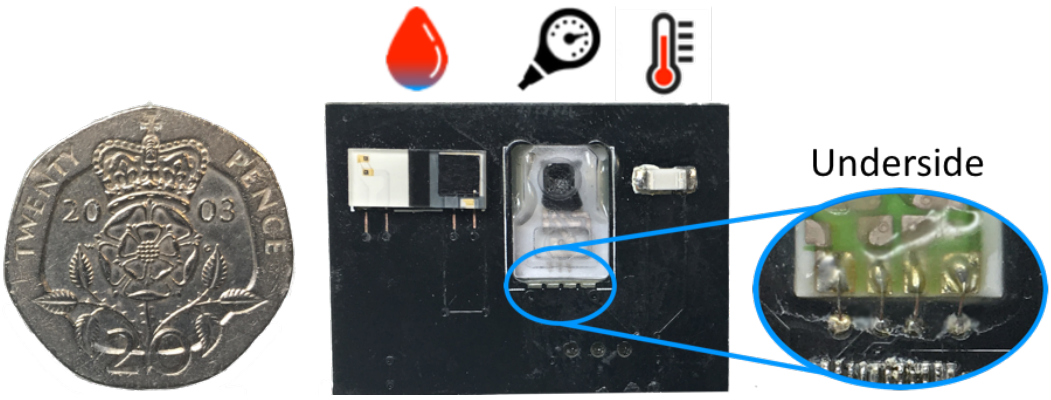


Figure 8.41: Positions of the sensors on the top PCB next to a 20p coin for scale. Note the hand-soldered connections to the pressure sensor on the underside of the board, sealed beneath a layer of epoxy to prevent accidental contact

Designing the Pipe

The internal diameter of the pipe was designed to match that of a normal renal artery at 6mm. However, the DCM03's 4.3mm diameter required the pipe to be designed with a D-shaped cross-section. I wrote a program in Matlab to determine the optimal diameter a circle missing a 4.9mm-long segment to ensure the cross-sectional area of the 'D' with 300μm lateral clearance for the DCM03 equalled that of a standard 6mm-diameter pipe.

Figure 8.42 below shows a transparent rendered image of the interface pipe beside an image of the pipe in profile. These demonstrate its internal recesses tailored to house the individual sensors and that the sensors all come to lie flush with the base of the 'D' section so as not to interfere with laminar flow through the pipe.

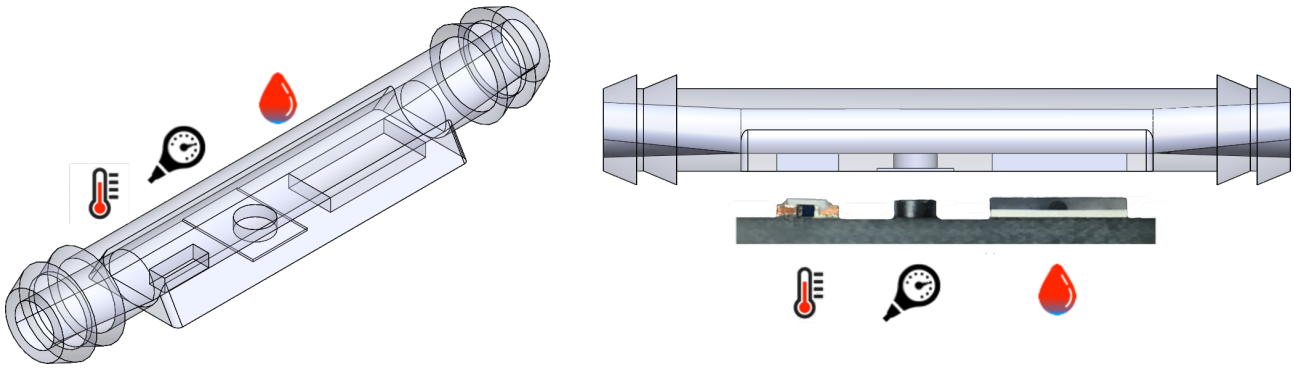


Figure 8.42: Left: Rendered image of the interface pipe demonstrating the recesses tailored to each sensor **Right:** A transparent profile image and side-view of the top PCB demonstrating how the sensors come to lie flush with the base of the internal 'D'-shaped lumen

The pipe was affixed to the top board with judicious use of epoxy to ensure that the internal recesses were watertight without any overspill into the internal lumen, and then painted

black with 3 layers of water-based paint to prevent external light interfering with the oxygen saturation sensor. This final assembly is shown in Figure 8.43 below:

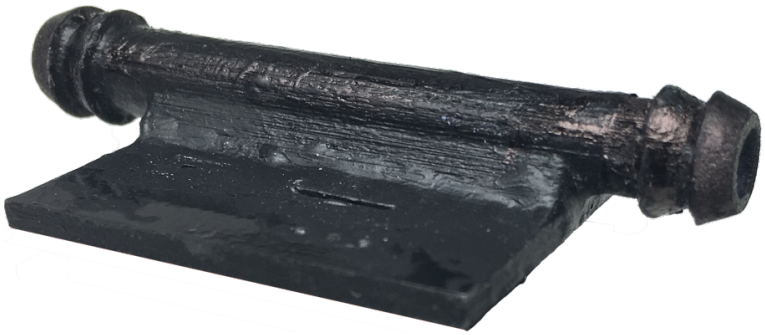


Figure 8.43: Photograph of the fully-assembled disposable top layer of the SmartPipe painted black to prevent external light from interfering with the oxygen saturation sensor

Chapter 9

The Sensor Conditioning and Digitisation (Middle) Layer of the SmartPipe

This Chapter describes the design and implementation of the analog signal conditioning systems for the temperature, pressure and oxygen saturation sensors integrated into the SmartPipe. The design employed a pair of low-cost, fully-integrated devices from Texas Instruments – the AFE4490 and LMP90080.

9.1 Chapter Overview

This Chapter synthesises the work of the previous Chapters in this part of the Thesis, particularly the selection of the SmartPipe's sensors for the top layer.

The design and layout of the middle-layer board will be discussed first, followed by the implementation of the temperature and pressure sensors with the LMP90080, and the oxygen saturation sensor based on the AFE4490. The Chapter will finish with power savings achieved through software modifications, and conclusions regarding the development of the SmartPipe.

9.2 Design and Layout of the Middle Layer Board

The middle PCB plays the role of interfacing between the analog sensors of the top board and the digital calculation and output stage of the bottom board. It contains only the two devices highlighted in the diagram of the SmartPipe shown below in Figure 9.1.

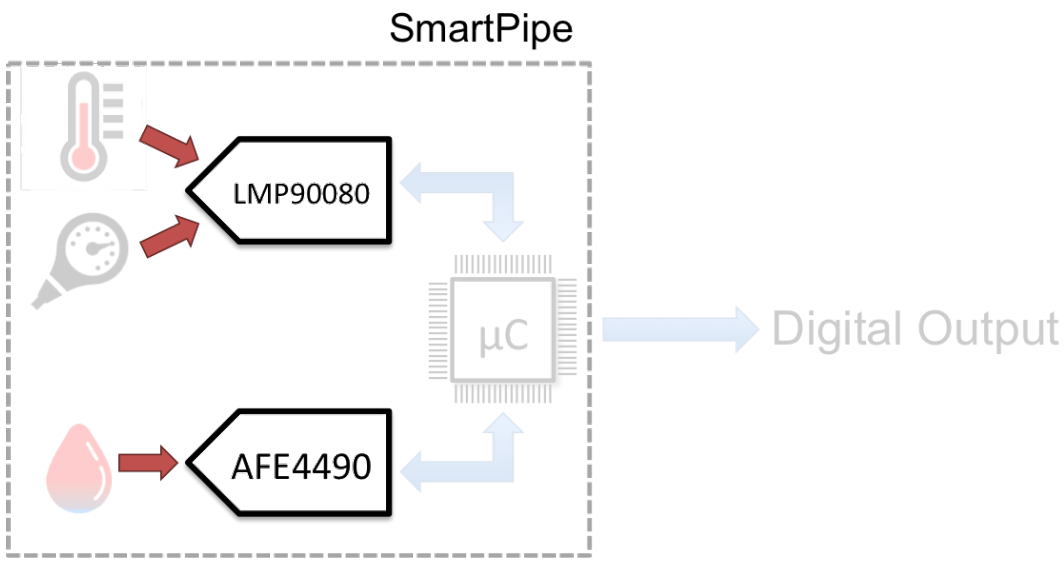


Figure 9.1: Role of the LMP90080 & AFE4490 in the SmartPipe. Analog signals are indicated by red arrows, with digital signals shown in blue

Separation of Analog and Digital Signals

To minimise noise and distortion in the signals from the front-end sensors, the analog and digital components were separated as widely as possible. Applying the principles of good board layout from Section 6.4 on page 180, the sensitive analog traces were routed on the

surface of the board, and kept as short as possible. Most digital signals and current-carrying traces were routed beneath the board and could take more tortuous paths.

A ground plane sat beneath the upper analog layer to prevent current loops and reduce stray EM noise as far as possible. The power lines were routed in a point-to-point fashion avoiding the analog section, rather than using a power plane. Figure 9.2 below demonstrates the mixed-signal layout on these three layers. I routed all traces by hand.

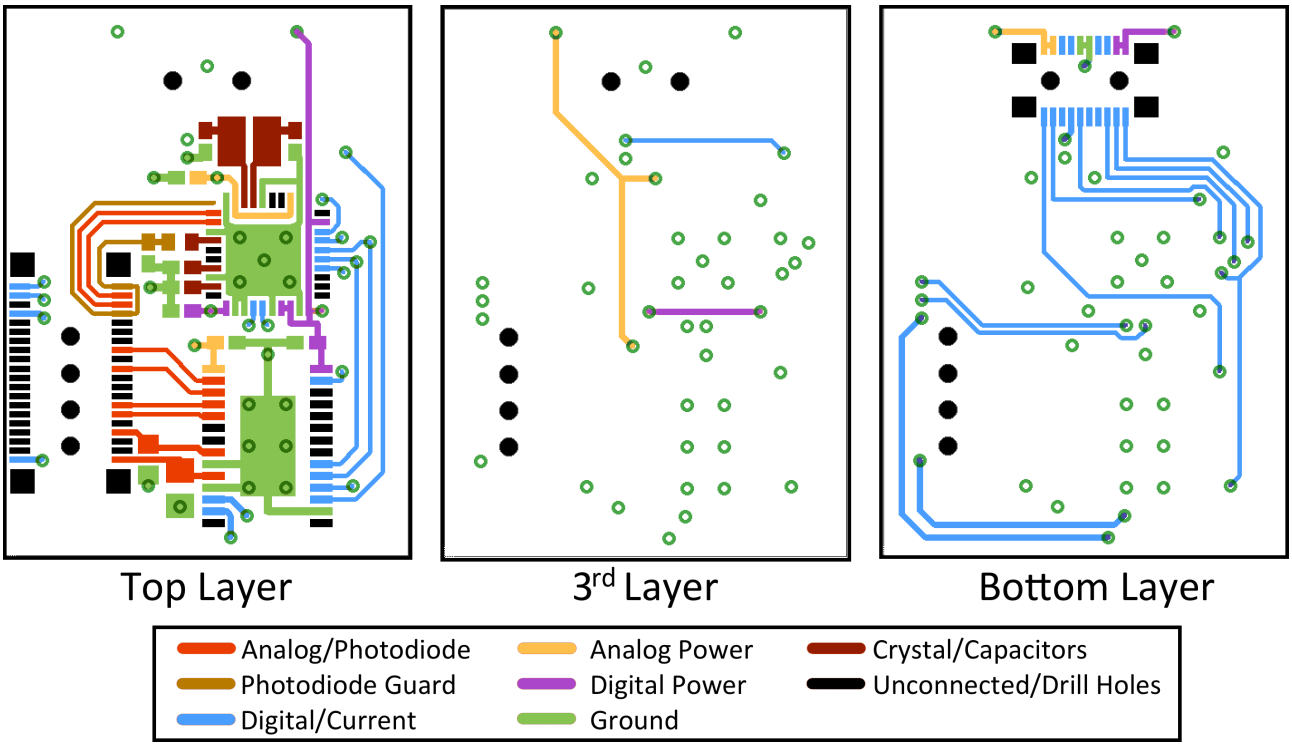


Figure 9.2: Top-down view of the internal layout of the middle layer mixed-signal PCB. The LMP90080 (elongated green pad), AFE4490 (square green pad) and middle to top board connector (parallel rows of black pads) sit on the top layer

This design resulted in the board shown in Figure 9.3 below.

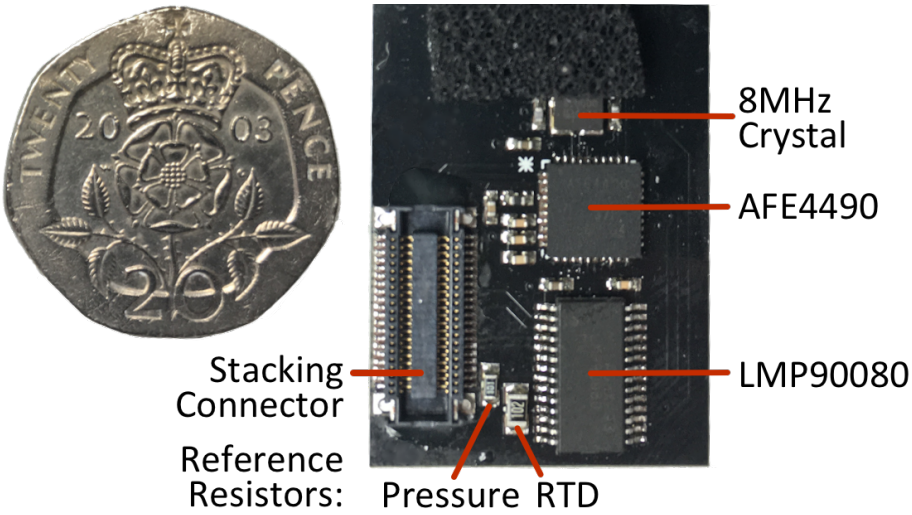


Figure 9.3: Image of the finalised mixed signal PCB forming the middle layer of the SmartPipe, alongside a 20p coin for scale. The foam pad at the top of the picture keeps the top board lying flat

Board-to-Board Connector

The design concept outlined at the beginning of Chapter 5 was for the SmartPipe to function as a reusable device in order to keep operating costs low and to commoditise warm organ perfusion technology. To this end, the top layer of the board discussed in Chapter 8 was designed to be disposable, meaning that the more costly electronics of the middle and bottom boards could be recycled between uses.

As such, the top board carries only the sensors, with all power supply and signal conditioning by the middle board. The signals carried by the middle-to-top connector, a 40-pin, 500μm SlimStack™ receptacle[1] from Molex[2], are shown in Figure 9.4 below. This mates with a paired 4-pin header[3] on the underside of the top board.

[1] 54363-0489
[2] Molex Inc., Lisle, Illinois, USA
[3] 55650-0488

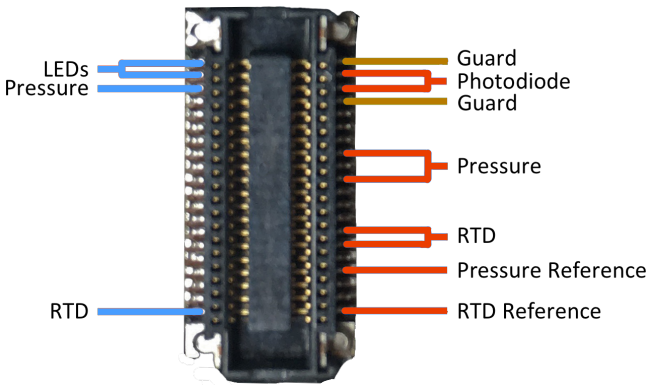


Figure 9.4: Signals carried by the board-to-board connector between the middle and top PCBs. Red lines carry analog signals, blue lines carry currents, brown indicates the photodiode guard line

The connection from the middle board to the digital board was shown previously in Figure 7.2 on page 194.

9.3 Digitising the Temperature and Pressure Signals

Both the temperature and pressure sensors[1][2] selected for the SmartPipe are resistance-based analog devices producing fully-differential signals. They can be excited by current sources, and would benefit from ratiometric referencing with different reference values.

The LMP90080[138] is an *analog front end* (AFE) device from Texas Instruments designed to greatly simplify the design process for electronics engineers by integrating all components required for analog signal generation and conditioning into one 10mm x 5mm surface-mount package, as shown in Figures 9.5 and 9.6 below.

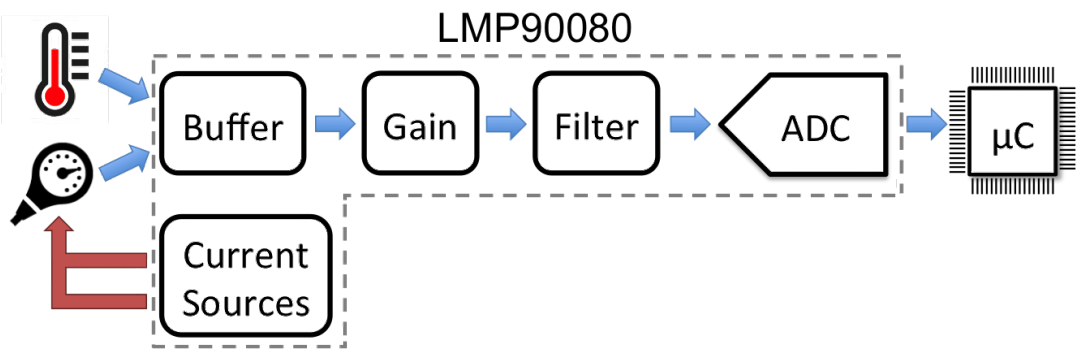


Figure 9.5: Details of the analog sensor conditioning functionality within the LMP90080

[1] A 1206 SMD Chip Pt100 RTD from Innovative Sensor Technology IST AG, Ebnat-Kappel, Switzerland
[2] The 1620-4-N from Measurement Specialties, TE Connectivity Ltd., Schaffhausen, Switzerland

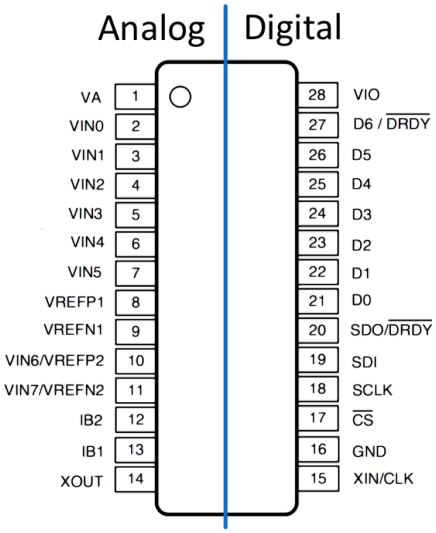


Figure 9.6: LMP90080 pin layout demonstrating the separation of the analog and digital pins

The LMP90080's multiple inputs feed into impedance matching buffers followed by low-noise, high-accuracy and low-drift gain amplification for signal conditioning. After conditioning, signals are piped into a 15-bit accurate $\Delta\Sigma$ ADC (16-bit 'claimed' accuracy actually equals 15 data bits + 1 sign bit) which also integrates a user-controlled fixed gain amplifier, a digital power-line filter, and circuitry for continuous calibration and correction of the signal gain and zero offset. The LMP90080 also provides two digitally-controlled current sources and two ratiometric referencing circuits.

The LMP90080 is an SPI slave device, requiring setup and configuration by the ATMega328P host microcontroller. The LMP90080 provides a $\overline{\texttt{DRDY}}$ signal to inform the host microcontroller that the ADC conversion has been completed and that there is data waiting to be read out from the `ADC_DOUT` registers at addresses `0x1A` and `0x1B`. This was key to my implementation of an interrupt-driven software design as previously discussed on page 206 in Chapter 7.

Other key features included the degree of flexibility and customisation possible in selecting the input sources, data conversion rates, gain settings and the continuous background offset correction system.

Input Multiplexing

The LMP90080 has a multi-channel multiplexer at the input stage, which feeds signals into the buffers. This multiplexer has certain pre-set differential channels, but these are fully configurable by the host microcontroller by manipulating the data in the relevant `CHx_INPUTCN` registers. The final embodiment of the SmartPipe used the default settings of Channel 0 ($V_{IN}0 - V_{IN}1$) for the pressure, and Channel 1 ($V_{IN}2 - V_{IN}3$) for the RTD.

Input Buffering and Amplification

The LMP90080 is the only member of the LMP900XX product family with 4 fully-differential channels with a layout allowing for good physical separation of the channels on the final circuit board. The input buffers are chopper-stabilised in order to cancel any offset drift, improving the accuracy and reproducibility of measurements. This leads to a typical offset error of only 11nV at a gain of $\times 128$[1] and a temperature drift of less than 0.2nV/°C.

Following the selection of input combinations via the multiplexer and buffering of the incoming signal, the voltage passes through a precision Fixed-Gain Amplifier (FGA) with binary gain steps from $\times 1 - \times 16$. For gain values >16, the system engages the buffers and the Programmable Gain Amplifier of the ADC, increasing the range of gains available from $\times 16 - \times 128$, again in binary steps. Thus a total gain of x64 uses a combination of the FGA at $\times 16$ and the PGA at $\times 4$.

User–Controlled Current Sources

Texas Instruments has integrated two precision, digitally-controllable current sources into the LMP90080. These may be programmed by the host microcontroller to output between 100μA and 1000μA in steps of 100μA. Unfortunately the single assignment controls the magnitude of both sources simultaneously, but this is a minor inconvenience.

The two current sources were programmed to deliver 500μA in order to balance the improvement in accuracy and sensitivity from a larger magnitude voltage drop across the sensors against the power drawn by two continuous sources in a system powered by a low-capacity 110mAh LiPo battery (page 196).

Ratiometric Referencing

For a detailed discussion of ratiometric referencing, please see Section 8.3 on page 233. With specific regard to the LMP90080, this device provides two independently selectable ratiometric referencing inputs, such that two different precision resistors can be used and assigned to provide the reference for individual channels or pairs of channels in differential mode. The selection of resistor values for each signal type is discussed later under the circuit designs for the temperature and pressure sensor on pages 294 and 297, respectively.

[1] Maximum 700nV

Configurable Data Conversion Modes

The LMP90080 can perform data conversions in two different modes – *single-shot* and *scanning*. In the *single-shot* mode, the LMP90080 wakes from sleep, samples and converts a single channel and makes this data available to the host microcontroller before returning to sleep. This mode requires the host microcontroller to manage the sampling rate of the system, and is really only useful for sub-Hz sampling or for fast host microcontrollers with accurate real-time clocks – not the ATMega328P.

The SmartPipe operated in the *scanning* conversion mode, performing continuous interlaced conversions of signals on Channels 0 and 1 at the sampling rates programmed into the `ODR_SEL` bits of the individual `CHx_CONFIG` registers.

Configurable Data Rates

By manipulating the `ODR_SEL` bits (bits `[6:4]`) of the `CHx_CONFIG` registers of a given channel, the user may select from a variety of sampling rates between 1.6775 Hz and 214.65 Hz. Scanning multiple channels at different data rates does slightly reduce the actual sampling rate for each channel. While the $\Delta\Sigma$ ADC can settle within a single cycle, the combination of higher sampling rates and higher gain does reduce the 15-bit noise-free performance of the ADC as shown in Table 9.1 below.

SPS	Gain							
	x1	x2	x4	x8	x16	x32	x64	x128
1.68	15	15	15	15	15	15	15	14.5
3.36	15	15	15	15	15	15	15	13.5
6.71	15	15	15	14.5	15	15	14	13.5
13.42	15	15	14.5	14	15	14.5	14	13
26.83	15	15	15	15	15	15	14.5	14
53.66	15	15	15	15	15	15	14	13.5
107.32	15	15	15	14.5	15	14.5	13.5	13
214.65	15	15	14.5	14	15	14	13.5	12.5

Table 9.1: Heat map representation of the ENOB performance of the 15-bit unipolar LMP90080 for specified combinations of gain and sampling rate (SPS). Data from [138]

As such, it is important to select a 'sweet-spot' of performance, finding a compromise between the gain magnitude and the desired sampling rate for a specific signal type, as first presented in Table 5.1 on page 166 in Chapter 5.

Continuous Background Calibration of Gain and Offset

Even with the use of a chopper buffer and precision gain amplifiers, it is possible for the system to develop non-zero offset drift and gain errors over time by compounding small errors along the analog signal chain.

The LMP90080 therefore has the ability to apply correction factors to the gain and offset of the analog input chain during the digital post-processing stage of ADC's output.

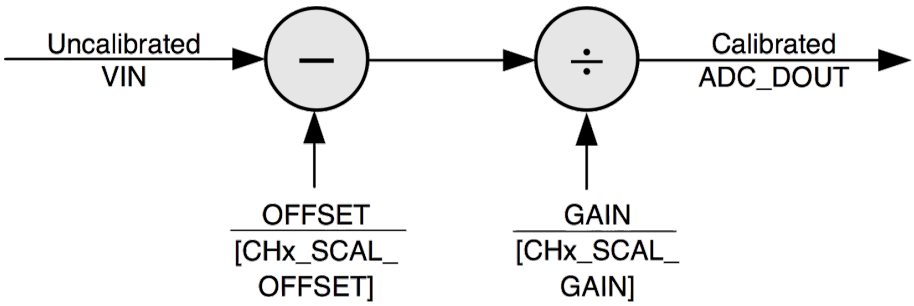


Figure 9.7: LMP90080 gain and offset correction. From page 22 of [138]

I determined that the optimal method for using the autocalibration mechanisms of the LMP90080 was to first place the system into **Estimation** mode for both the gain and offset, followed by a *single conversion* in **Correction** mode to update the channel's gain & offset registers. The system would then return to **Estimation** mode and continue operating with the new correction factors, because as I discovered, operating the system permanently in **Correction** mode produced very unexpectedly noisy outputs.

Summary of the Important Control Registers for the LMP90080

Register	Address	Function
`BGCALCN`	`0x10`	Configures background calibration and correction mode
`ADC_AUXCN`	`0x12`	Controls the current sources
`SCALCN`	`0x17`	Activates system offset and gain coefficient determination
`CH_SCAN`	`0x1F`	Scan mode selection and channels
`CHx_CONFIG`	Various	Controls channel sampling rate, gain, and buffer activation
`CHx_INPUTCN`	Various	Select channel inputs and reference

Table 9.2: Summary of the LMP90080's important control registers

9.4 RTD Circuit Design and Layout

The detection circuit for the RTD sensor is based on the 4–wire circuit shown on page 232, and takes advantage of ratiometric referencing.

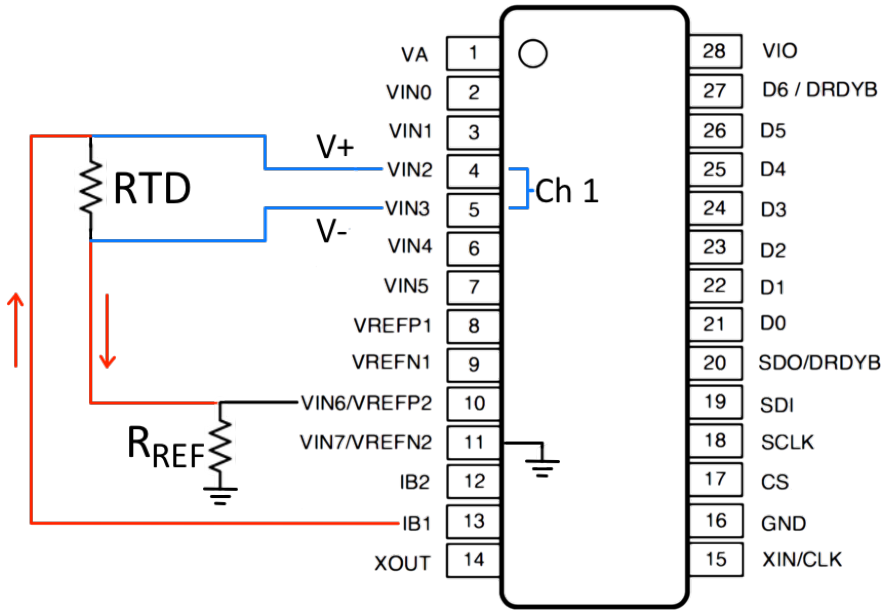


Figure 9.8: 4-wire Kelvin connection circuit between the LMP90080 the RTD, with the current path in red leading to the ratiometric referencing resistor, and the voltage sensing path in blue

Selection of the Gain and R_{REF} Values

Per the discussion on page 291, the current sources were programmed to deliver 500μA to balance the increased signal quality against an increased power draw.

This current value results in the following sensor voltages at 0°C and 40°C:

$$V_{0°C} = 500\mu A \times 100.00\Omega = 50mV$$
$$V_{40°C} = 500\mu A \times 115.54\Omega = 57.77mV$$

In order to maximise the stability of the reference voltage, I selected a <£10, 1kΩ resistor with 0.01% tolerance in an 0805 surface-mount package. Lower value resistors were available with 0.01% tolerances, but these would have resulted in a reference voltage ($I_{SOURCE} \cdot R_{REF}$) at or below the minimum signal expected at 0°C.

It was possible to select a binary-step gain magnitude to maximise the ADC's input range from the relationship between the magnitudes of the current source, reference resistor and gain as follows:

$$\text{Full-Scale Input} = \pm\frac{I_{SOURCE} \cdot R_{REF}}{\text{Gain}}$$
$$= \frac{500\mu A \cdot 1k\Omega}{8} = 64mV$$

The 57.77mV signal at 40°C represents 90.2% of this range, indicating good remaining signal headroom. A rough estimation of the resolution of the 15-bit unipolar signal is as follows:

$$\frac{64\text{mV}}{(2^{15}-1)} \simeq 0.00195\ {}^{\text{mV}}/_{\text{LSB}}$$

$$\frac{7.77\text{mV}}{40°\text{C}} \simeq 0.1942\ {}^{\text{mV}}/_{°\text{C}}$$

$$\approx 100 \text{ LSB per °C}$$

Board Layout of the Temperature Sensor Circuit

Figure 9.9 below illustrates the RTD sensor circuit as laid out across the top and middle PCBs. These two layers have been placed side-by-side, with the connections clearly marked. The dotted lines indicate connections between the layers completed via the board-to-board stacking connector:

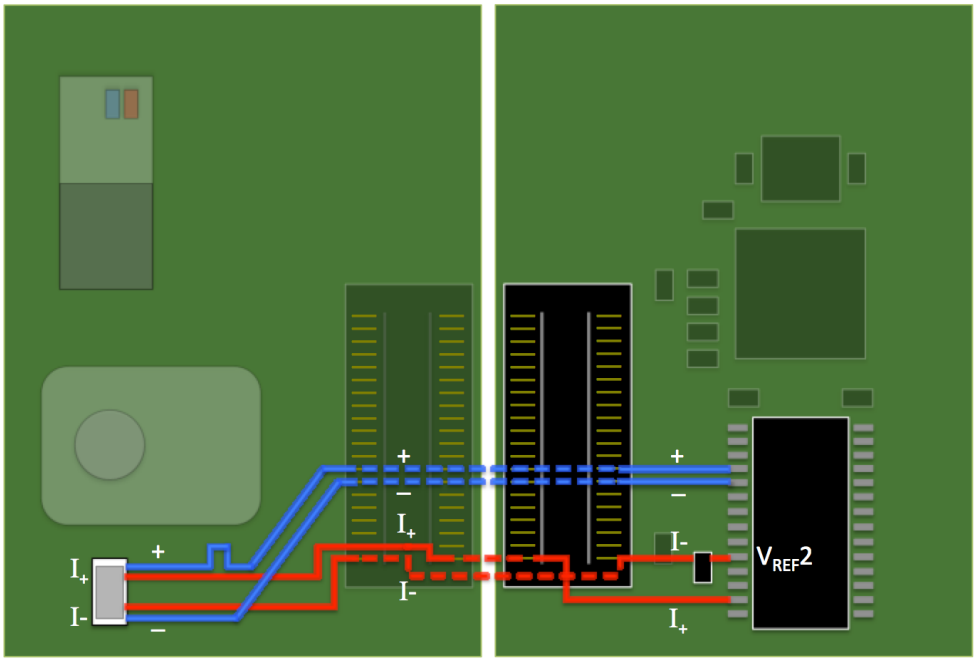


Figure 9.9: Layout of the 4-wire circuit connecting the RTD on the top layer of the SmartPipe to the LMP90080 on the middle layer, with current in red and voltage in blue. Note the top layer is shown from the underside, so that the image 'opens' like a book down the centre

In this layout the current carrying lines are shown in red, and the voltage sensing lines in blue. The current source for the RTD is separate from that for the pressure sensor, and

flows back to a precision 1kΩ, 0.01% resistor whose voltage is sensed at $V_{REF}2$. The voltage drop across the RTD feeds into Channel 1 ($V_{IN}2 - V_{IN}3$).

Note the following feature in the blue 'V+' line:

Figure 9.10: Detail of path-length matching

This additional trace length was inserted in order to match the resistances of the voltage sensing lines, minimising any potential mismatch error were any current to flow in them. The path lengths of the current carrying lines did not need to be matched in the same way.

9.5 Pressure Sensor Circuit Design and Layout

The detection circuit for the piezoresistive MEMS pressure sensor is based on a current-driven full-bridge connection illustrated on page 271, and takes advantage of ratiometric referencing.

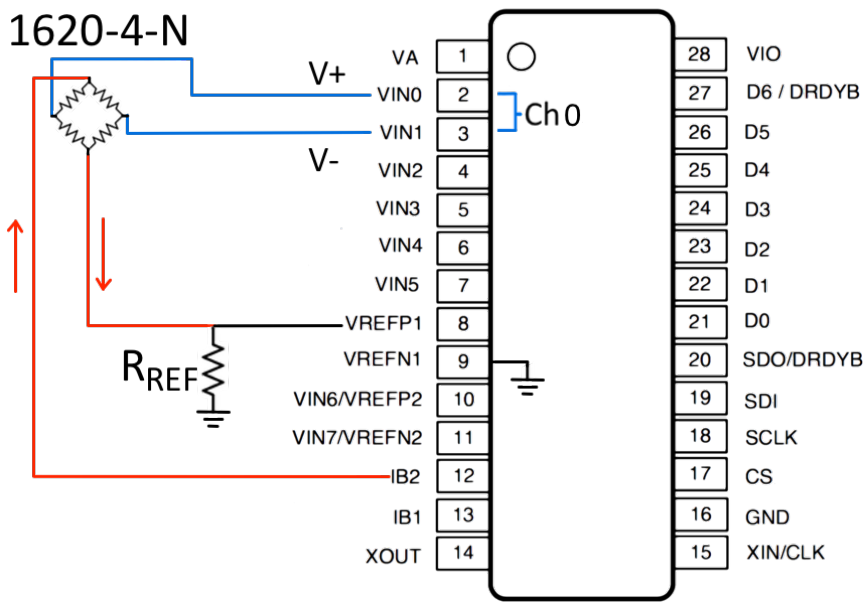


Figure 9.11: Current-mode bridge sensor circuit for connecting the pressure sensor to the LMP90080, with the current path in red leading to the ratiometric referencing resistor, and the voltage sensing path in blue

Selection of the Gain and R_{REF} Values

As with the RTD circuit, the current sources were programmed to deliver 500μA to balance the increased signal magnitude against an increased power draw.

The measured change in the resistance of the bridge was only 1Ω at 100mmHg, just 10mΩ/mmHg, meaning that the voltage across the sensor would only be 1mV at 200mmHg full-scale, far lower than that of the RTD.

For this reason it was necessary to use a much larger 64$\times$ input gain setting and lower R_{REF} value of 100Ω with 0.1% tolerance, the maximum tolerance available through commercial suppliers while remaining <£10/unit. This was in a smaller 0603 surface-mount package in order to lay out both reference resistors as close to the LMP90080 as possible.

Note that the precise value of the resistor was less important for the design of the pressure sensor than for the RTD because of the need for calibration.

Board Layout of the Pressure Sensor Circuit

Figure 9.12 below illustrates the circuit laid out across the top and middle layers in order to implement the pressure sensor:

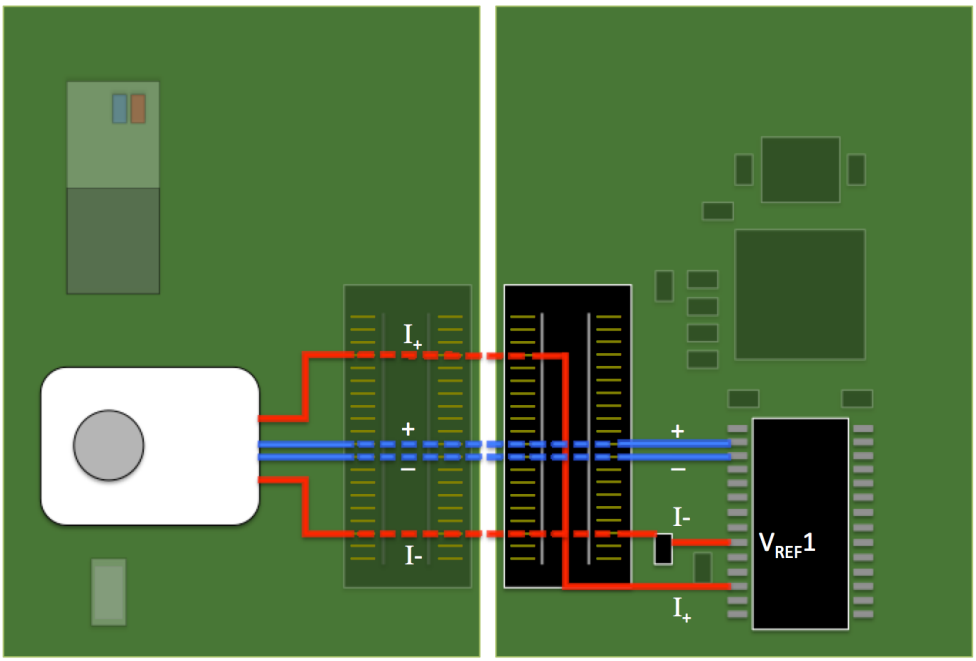


Figure 9.12: Layout of the connections between the pressure sensor on the top layer of the SmartPipe and the LMP90080 on the middle layer, with current in red and voltage in blue. Note the top layer is shown from the underside, so that the image 'opens' like a book down the centre

This diagram illustrates the current traces in red and the voltage traces in blue. The return path of the current limb (I-) passes to the precision 100 Ω, 0.1% resistor whose voltage is sensed at $V_{REF}1$.

The voltage drop across the pressure sensing bridge feeds directly into Channel 0 ($V_{IN}0$ – $V_{IN}1$), keeping the trace lengths as short and equal as possible in order to maximise noise immunity.

9.6 Testing the Temperature and Pressure Sensors

To briefly recap, both sensors were designed to take advantage of the 15-bit differential signalling capabilities of the LMP90080, with current stimulation and ratiometric referencing through different resistors.

Testing the Temperature Sensor

With the gain set at x8 to maximise the sensitivity of the RTD across the biological sensing range, the heat map of ENOB values in Table 9.1 on page 292 shows that only four of the potential sampling rates provide a full 15-bit result.

For this reason I programmed the LMP90080 to sample the temperature channel at a rate of 26.83 SPS, far exceeding the required bandwidth requirements first specified in Section 5.5.1 in Chapter 5, but potentially allowing for a digital averaging filter to be implemented later in software if required.

Calculating the Temperature

With the confidence of the LMP90080's auto-zeroing front end buffers, calibration and gain correction mechanisms, the effects of any offset and gain nonlinearity can be largely discounted. The accuracy of the digital code therefore derives solely from the accuracy of the reference resistor, specified at 0.01%.

With a ratiometric referencing configuration (page 233), the digital code produced by the LMP90080's ADC simplifies down to a ratio of two references, multiplied by the gain of the front-end amplifier. The magnitude of the current through each sensor is in fact irrelevant.

$$\mathrm{ADC_{OUT}} = \mathrm{GAIN} \cdot \left(\frac{\mathrm{R_{RTD}}}{\mathrm{R_{REF}}}\right) \cdot (2^{15} - 1)$$

This allows for a simple single-step multiplication within the ATMega328P to derive the resistance value of the RTD from the ADC's output code, which can then be used with the Callendar-Van Dusen equation (page 227) to calculate the temperature:

$$\mathrm{R_{RTD}} = \mathrm{ADC_{OUT}} \cdot \left(\frac{1000\Omega}{8 \cdot (2^{15} - 1)}\right)$$

Results of Testing

The RTD was affixed into the wall of a plastic beaker full of water atop a stirring hot plate and compared against a Comark N9094 digital thermometer. A short 4-wire connection was run between the RTD and an LMP90080 mounted on a prototype board connected to an Arduino. Because the DAC was unable to accurately display results of greater than 11 bits of resolution (Figure 6.9 on page 190), the temperature calculated on the Arduino

was displayed as numerical data directly over the serial port. Temperatures were manually recorded over several runs to gather most of the integer range between 0°C – 45°C.

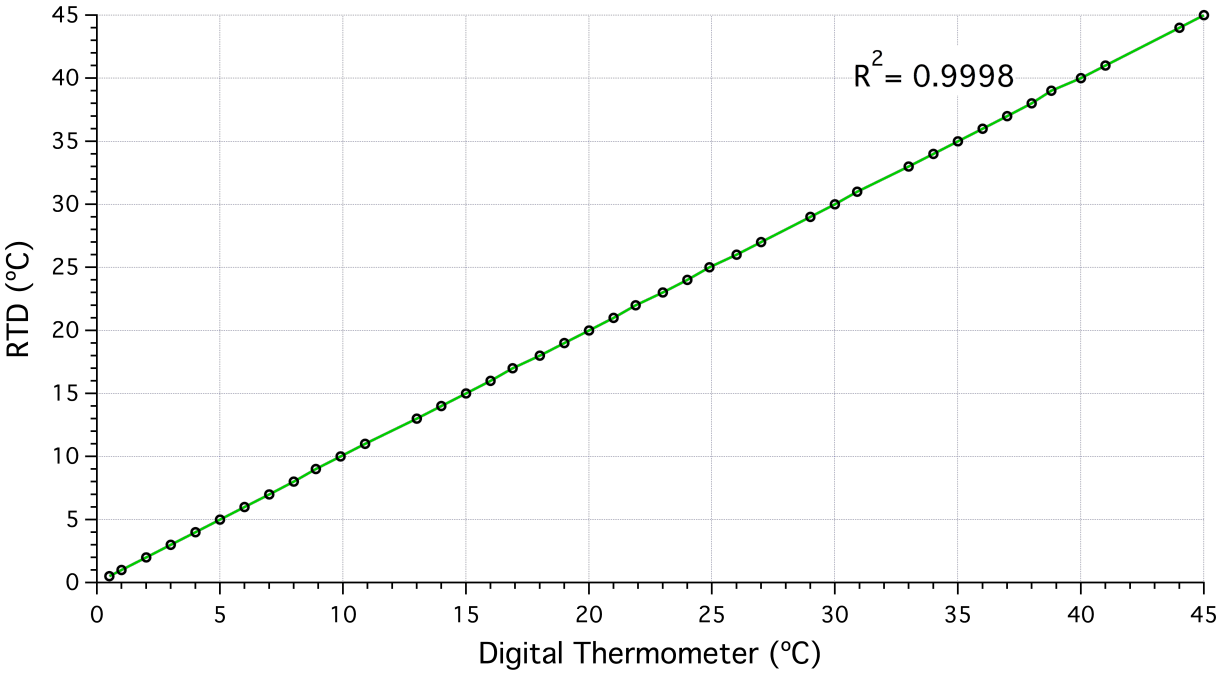


Figure 9.13: RTD tested against the Comark N9094 digital thermometer

The result shown in Figure 9.13 is not surprising, given that this was a test of one highly accurate digital thermometer against another.

Interestingly, the Comark N9094 uses a T-type thermocouple probe with a documented accuracy of ±0.2°C, better than that of the Class A RTD of ±0.23°C across the entire range of 0°C – 40°C. This thermocouple-based sensor must derive its readings from a fully-specified look-up table (LUT) in order to compensate for the many sources of inaccuracy described in Section 8.2 on page 219.

The SmartPipe temperature sensor performed as expected over the 0°C – 40°C biological temperature range and with a bandwidth in excess of 1Hz.

Testing the Pressure Sensor

The LMP90080 was programmed to sample the pressure sensor at its maximum rate of 214.64 SPS to meet the requirements of a biological pressure sensor system as discussed on page 161 in Chapter 5. Referring back to Table 9.1 on page 292, we can see that this combination of sampling rate and 64× gain should provide us with a maximum 13.5 bits of noise-free digitised data from the sensor. Gain and offset corrections were applied as described on page 293.

The sensor was calibrated in the SmartPipe against the dial gauge of a Welch Allyn medical sphygmomanometer[1], accurate to the nearest 2mmHg. Its output was relayed directly through the DAC into LabChart, from where the calibration data were collected.

The image below demonstrates the linearity of the response (n=5) and the very narrow standard deviations across a pressure range of 0mmHg – 180mmHg (±1–4mV for all readings except for ±18mV at 180mmHg). The insert reproduces the raw data from one of the step-wise pressure tests.

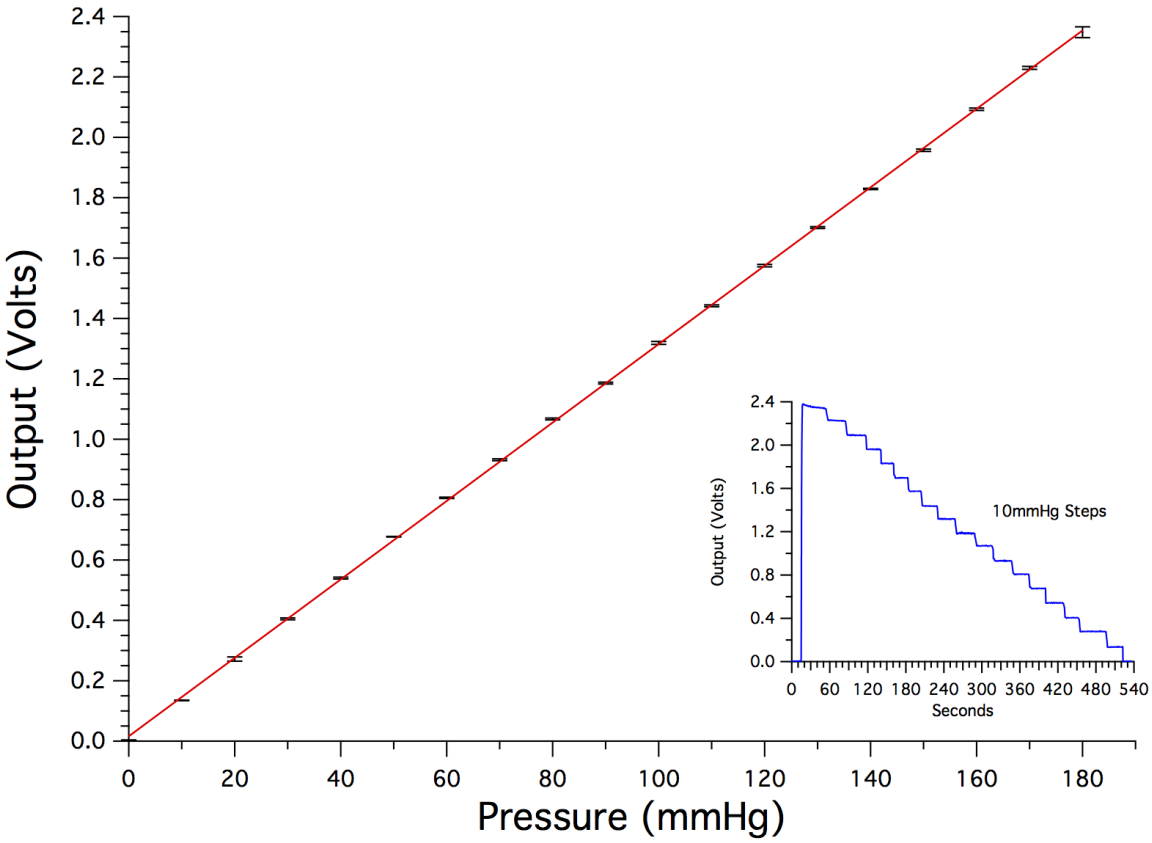


Figure 9.14: Linearity of the pressure sensor under test conditions. Insert: Raw trace of one test cycle

The calculated Limit of Detection (LOD) is equal to:

[1]Model Number: SN151022232447

$$
\begin{aligned}
f(\text{LOD}) &= \frac{3 \times \text{SD}_0}{\text{Slope (mV/mmHg)}} \\
&= \frac{3\text{mV}}{13\text{mV/mmHg}} \\
&= 0.23\text{mmHg}
\end{aligned}
$$

The following graph demonstrates the deviation of each step change from ideal.

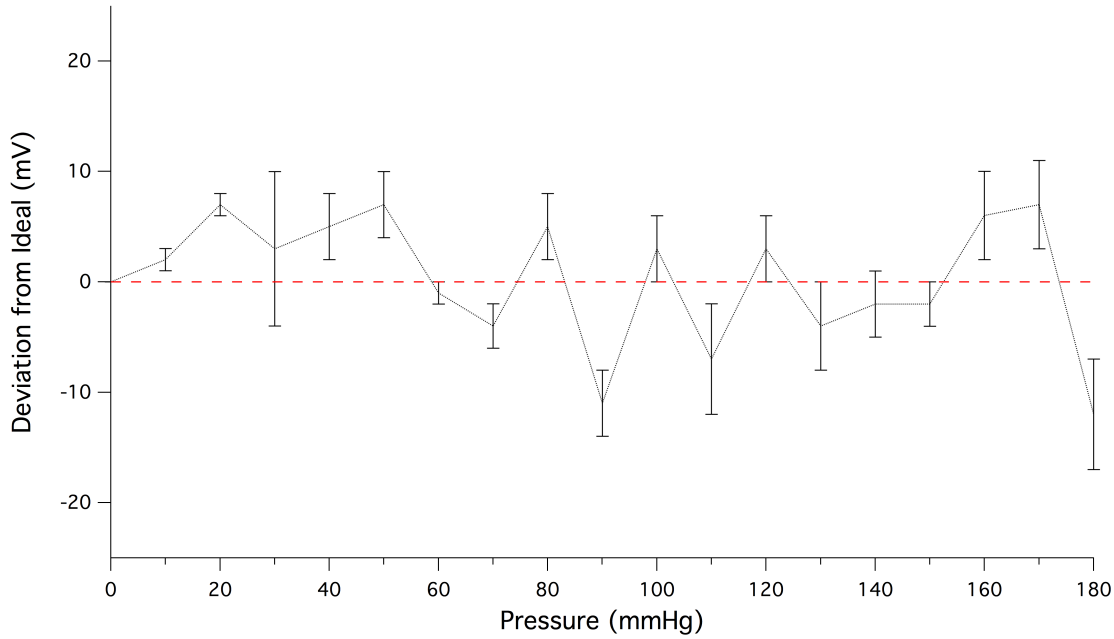


Figure 9.15: Deviation of the pressure sensor response from ideal

It is clear that the greatest deviations from ideal occur at 90mmHg and 180mmHg. This is perhaps related to the internal spring mechanism of the medical sphygmomanometer against which the pressure sensors were calibrated, rather than a uniform defect in the 1620-4-N pressure sensors, but it is not possible to determine this without further cross-calibration against a second standard.

Note that the noise level of 3mV_{RMS} gives us an ENOB for this sensor of ≈8.6 bits rather than the projected maximum of 13.5. This undoubtedly arises from a degree of degradation in the DAC→ADC stage (shown in Figure 6.9 on page 190), as well as the inaccuracies in the sensor described by the product's data sheet. Nevertheless, the SmartPipe's pressure sensor system is more than capable of meeting the design goals of detecting pressure changes over a range of 0mmHg – 180mmHg with 2mmHg resolution and 200Hz bandwidth.

Live Re-Scaling and Data Conversion

As per the discussion on page 184, at a later stage in software development during full system integration, the temperature was calculated and re-scaled directly on the SmartPipe to overcome the 11-bit limitation of the DAC-PowerLab system. Figure 9.16 below shows the raw voltage feed from the PowerLab on the upper channel, and the simultaneous display of temperature to the nearest 0.1°C.

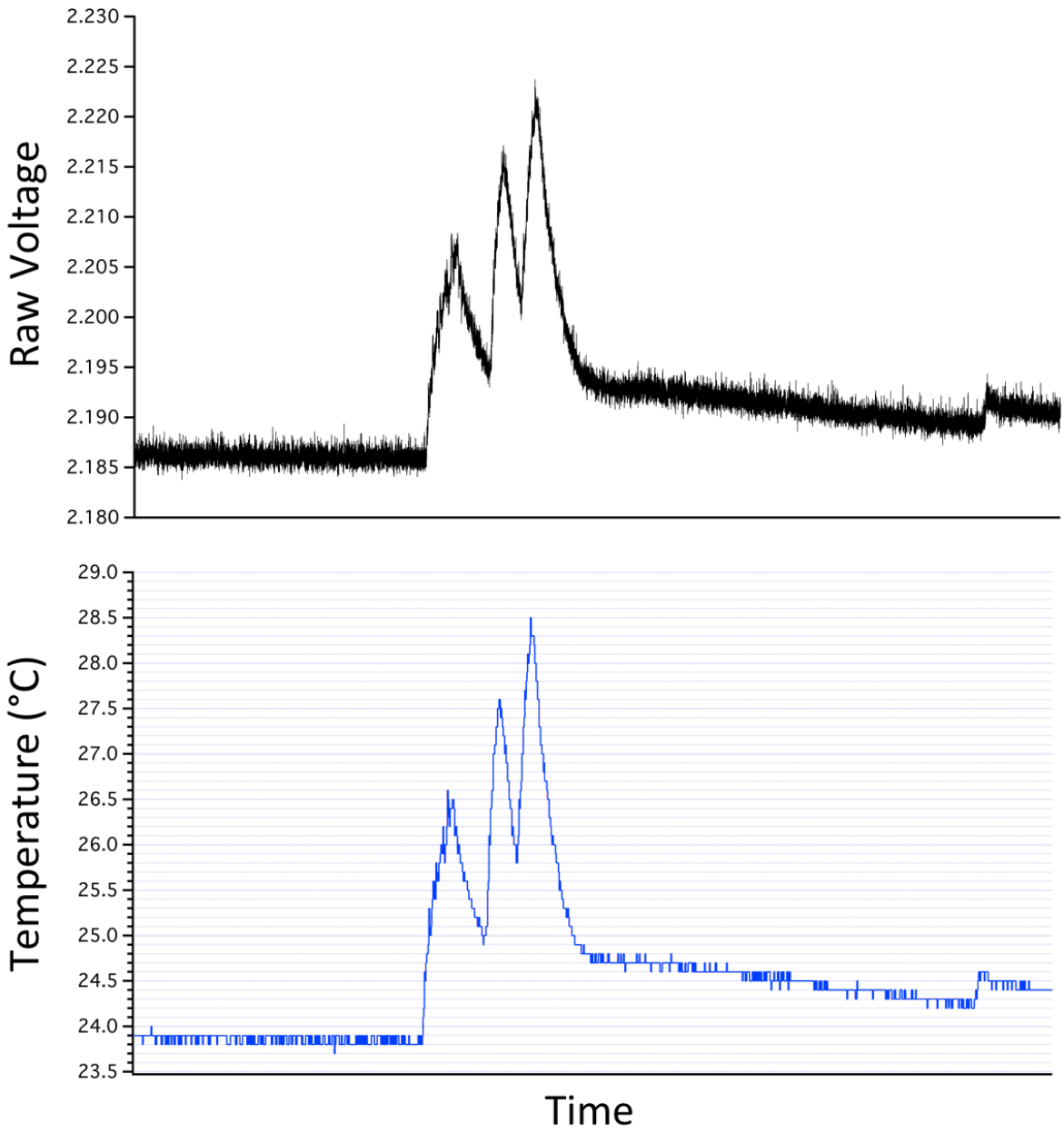


Figure 9.16: Live conversion and re-scaling of the RTD data into temperatures with 0.1°C resolution

This data demonstrates quantisation noise which is clearest at the beginning of the feed where the true signal value lies between 23.8°C – 23.9°C. This is unfortunately due to the

loss of resolution intended to overcome the noise of the original signal through the use of a low-resolution look-up table (LUT) (page 317). Here the software algorithmically determines the nearest match to the true input value, which means that values half-way between distinct steps (*e.g.* 23.85°C) will have equal probability of resulting in the output of neighbouring values in the table, thus the appearance of 'flip-flopping' between states.

There are a few ways that this could be improved such as the use of software averaging routines which would lower the signal bandwidth and thus frequency of step transition, dithering the output data (adding artificial broad-band noise after processing in order to 'smooth' the appearance of the output), or improving the resolution of the look-up table by 1–2 bits. The optimal solution would of course be to improve the overall resolution of the data chain, so that these quantised transitions occur at a level of less than half the desired minimum step value.

9.7 Measuring the Oxygen Saturation

The use of the DCM03 sensor was intimately linked to the use of the AFE4490, a fully-integrated analog front end (AFE) device designed for the collection of intensity data from paired LEDs.

Historically, oxygen saturation measurements have required a combination of components including LED power regulation to ensure stable intensity, a means of time-gating the LED to effectively multiplex the two wavelengths onto one photodiode, a transimpedance stage to measure the signal from the photodiode, and finally appropriate gain and digitisation circuitry. The AFE4490 combines all of these components into an ultra-small 6mm×6mm×1mm surface mount device at a cost of <£15/unit, the internal configuration of which is shown below in Figure 9.17.

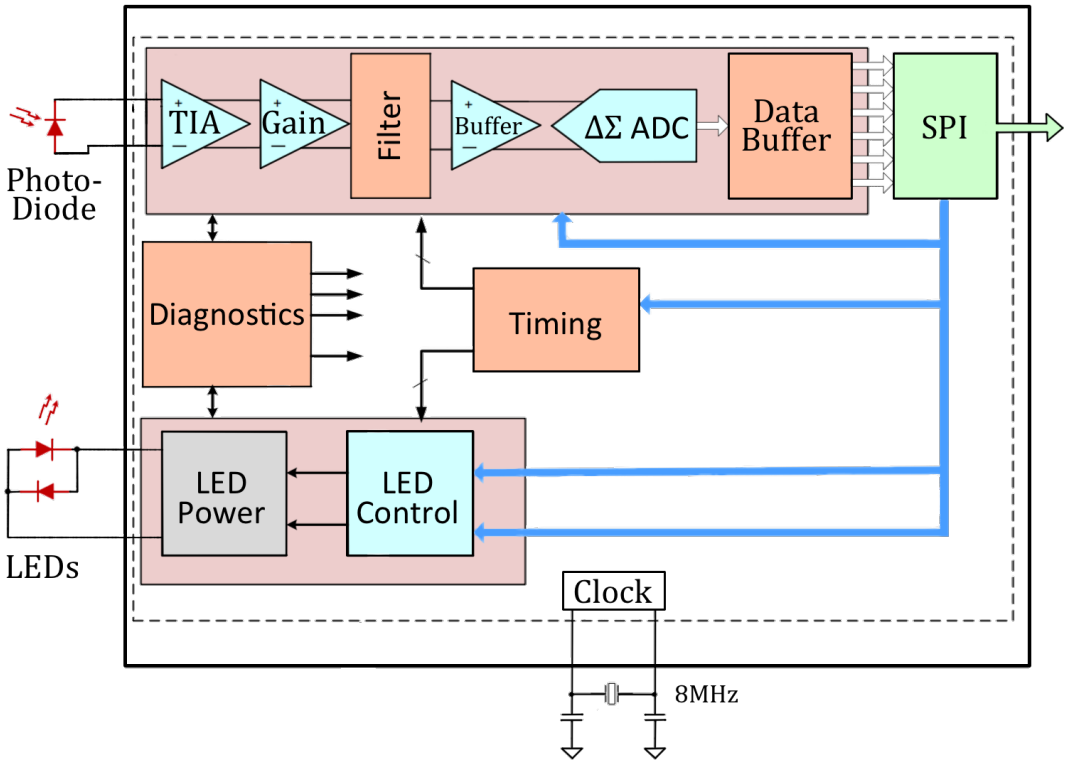


Figure 9.17: Internal detail of the AFE4490. From [139]

Some of the key features will be briefly considered here. Note that many of these systems are available to customisation by the system designer using the SPI interface.

LED Control

The AFE4490 contains an integrated LED drive system which can operate as either an H-bridge for antiparallel LEDs like the DCM03, or in a push-pull configuration for a common-cathode arrangement. The output intensity of each LED can be programmed either jointly, or independently. I elected to use the system with independent settings on each LED channel due to the degree of control this offered during experimentation.

This is managed by first informing the AFE4490 of the maximum full-scale LED voltage through register `CONTROL2`, chosen to be 50mA for the SmartPipe in the `CONTROL2` register due to the 3.3V digital supply voltage, followed by loading values into bits `[15:8]` and `[7:0]` for LEDs 1 and 2 into the `LEDCNTRL` register, according to the nearest integer value found by the following formula:

$$\text{LED Code} = \frac{256 \times \text{Output in mA}}{\text{Full-Scale Current}}$$

The determination of the optimal LED settings forms part of the results discussed later

from page 311.

Transimpedance Amplifier

The receiver front-end of the AFE4490 comprises a photodiode which feeds into a differential transimpedance amplifier (TIA) followed by a secondary gain amplifier. The TIA stage is again under user control. Here the user is able to set the value of the feedback resistor and stabilising capacitor either independently for each LED, or by using a single setting for both. Again, programming each channel for independent control offered me the greatest degree of flexibility during development.

The allows the user to select TIA feedback values of 10k Ω, 25kΩ, 50kΩ, 100kΩ, 250kΩ, 500kΩ or 1MΩ. The TIA feedback stabilising capacitor can be selected in 5pF steps from 5pF to 155pF. While there are formulæfor determining the optimum settings for the feedback value, the data sheet reports a resistor gain accuracy of only 7% as the device is intended to be used for ratiometric determinations rather than calculations of absolute signal magnitudes. Thus, a degree of experimentation with capacitance values to prevent excessive noise was required, whilst ensuring that the RC combination did not exceed $^{1}/_{10}{}^{\text{th}}$ of the sampling duration.

Secondary Gain Stage

The maximum output of the TIA stage is ±1V, with a maximum input into the ADC of ±1.2V. Operating the secondary gain stage does consume slightly more power, but it is able to maximise the use of the full range of the ADC and allow the use of smaller photocurrents and higher TIA gain values. The second stage has 5 selectable gain settings from 1×, 1.5×, 2×, 3× to 4×.

The AFE4490 is also able to feed an ambient signal cancellation current of between 1μA – 10μA into this stage to neutralise any background error signal arising from the photodiode's dark current. This was not used during the project as there was no appreciable change to the ADC's output during experimentation with the DAC-PowerLab system. However, this may have been a false negative result arising from signal-chain errors at this later stage rather than internal to the AFE4490 itself.

In a later development of the SmartPipe, the system could be programmed to select its own optimal LED output, TIA gain, secondary gain and ambient cancellation DAC settings through an automated feedback control loop. However, I wanted to keep the system under tighter control whilst exploring its behaviour and so did not code such a system during the project.

$\Delta\Sigma$ ADC

Following the secondary gain stage, the analog signal passes through a low-pass filter (500Hz or 1kHz, under user control) and buffer before feeding into an integrated 22-bit delta-sigma ADC with an internal 1.2V band-gap reference. The ADC requires an external $\geq$2.2μF capacitor to stabilise the reference voltage.

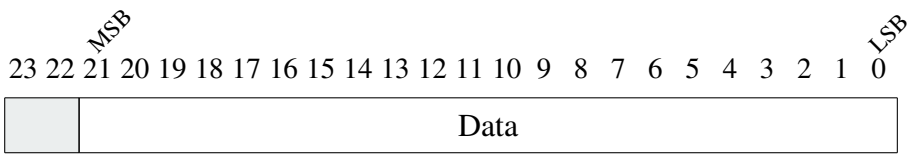
The ADC converts 4 signals derived from the input stage in sequence: Ambient LED2$\rightarrow$ LED1 $\rightarrow$ Ambient LED1 $\rightarrow$ LED2. See the discussion of **Timing Control** on page 309 for a more detailed discussion of this sequence.

After converting all 4 values, the AFE4490 loads these values into the digital output buffer and also calculates the difference between the signal and ambient values for two additional results. The AFE4490 informs the master device these six data are ready for collection over SPI by triggering a rising edge on the `DRDY` pin, triggering an interrupt on the bottom layer's ATMega328P.

Managing the ADC's output was one of the most difficult parts of this stage of the project and involved a lot of work on the AFE4490's software library, for reasons discussed below.

Reformatting the Data Word

The AFE4490 provides its results as 22-bit values in twos-complement format as shown below. The two bits beyond the MSB[1] are not used.



However, I found that the ATMega328P is not able to accept >16-bits into one address in a read sequence due to peculiarities of the interaction between the SPI routines and the internal addressing system of the 8-bit processor. Attempting a left-wise bit-shift of a 16-bit word into a 24-bit or 32-bit block to receive the 8 remaining LSBs over SPI causes loss of the 8 MSBs, leaving only the 16 LSBs at the original address.

I therefore had to reconstruct the 24-bit word in a piecemeal fashion, receiving the data into an 8-bit array with 3 separate SPI calls, and using a 32-bit temporary variable to reassemble the final 24-word byte from this array in 3 stages.

```

uint8_t rxBuffer[3] = {0};
uint32_t temp = 0;
```

[1]Most Significant Bit

```

rxBuffer[0] = SPI.transfer(0x00);
rxBuffer[1] = SPI.transfer(0x00);
rxBuffer[2] = SPI.transfer(0x00);

temp = rxBuffer[0];
temp = (temp << 8) | rxBuffer[1];
temp = (temp << 8) | rxBuffer[2];
```

Because the value of the photodiode will always be positive, despite the fully-differential receiving stage from the TIA through to the ADC, the twos-complement format of the data word needs to be changed into straight binary for manipulation and transmission to the unipolar DAC-PowerLab system for visualisation.

The full-range of a 22-bit twos-complement number is -2,097,152 to +2,097,151. Therefore, adding 2,097,152 (0x200000) to the result will shift the value into the unipolar range from 0 to 2^{22} or 0 to 4,194,304, as follows:

```
temp += 0x200000;
```

With the final data word assembled and occupying 22 bits of the 32-bit block, I had to right-shift the word back down into the 16 LSBs, perform a bit-masking operation and return this final 16-bit word to the main program. However, with a well-scaled combination of the TIA and secondary gain stage, the ADC should not be saturated and therefore the MSB of this 16-bit word will always be `0`, only occupying half the range of the DAC-PowerLab system. For this reason, I only right-shifted the 22-bit straight binary value down by 5 bits.

```
return ( (temp >> 5) & 0xFFFF);
```

Having this 1 LSB occupied by noise did not matter as the system is only specified for 13.5-bit accuracy (TIA and gain dependent), despite producing a 22-bit result.

SPI Peculiarities

The first peculiarity of the AFE4490's SPI communication subsystem was the need to manually switch the device into `READ` mode with each data collection. Each sequence required an 8-bit transmission to set the receiving address to `CONTROL0`, followed by 24 bits of data just to set the LSB of this register, switching the device into `READ` mode. Then followed the data collection itself, before repeating this 32-bit block but writing a final `0` to clear the `READ` bit.

The SPI subsystem implemented by the AFE4490 also had a non-obvious behaviour that I discovered during final system integration that required me to implement a work-around for use in a multi-slave configuration alongside the LMP90080.

SPI slave devices will usually tri-state their `MISO`, `MOSI` and `SCLK` pins automatically in hardware unless $\overline{\texttt{CS}}$ is pulled low to activate the device for communication. However, the AFE4490 required the tri-state to be set manually by the user in software, which had the side-effect of deactivating the AFE4490's `DRDY` pin. This meant that every call to the LMP90080 had to begin with a function to set the 10th bit in the `CONTROL2` register, and clear it again afterwards so that the AFE4490 could trigger its hardware interrupt. With 8 bits for the `CONTROL2` address, and 24 bits required to toggle the tri-state bit, this added 32 bits with each call.

Each of these unusual behaviours therefore added 64 bits of overhead to every SPI communication on the SmartPipe - 64 bits for the AFE4490 and 64 bits for the LMP90080, reducing the total data rate and the time that the ATMega328P could spend in a power-saving sleep mode. At a sampling rate of 200Hz for the pressure sensor and 62.5Hz for the AFE4490 (the lowest setting), this overhead alone equals $\geq$0.2% of the total power consumption of the SmartPipe[1].

Timing Control

The AFE4490 performed its LED activation, deactivation, sampling, and conversion in a strict sequence although customisable by the user through the 28 timing control registers. This sequence of simultaneous actions is shown in Table 9.3 below. The ADC is also triggered to reset between each block by its timing registers (not shown).

Period	1	2	3	4
LED		LED1		LED2
Sampling	aLED2	LED1	aLED1	LED2
Converting	LED2	aLED2	LED1	aLED1

Table 9.3: The start and end time of each part of the activity sequence of the AFE4490 is open to user modification but not the order. aLED = ambient, no LED active

The timing control circuitry derives its clock from the external 8MHz crystal, divided by 2. The resulting 4MHz is divided into 64,000 equal units of 250 nanoseconds, and the value programmed into each of the 16-bit timing registers sets the duration of each step. The total sampling rate of the system is determined by the repetition rate of the sequence in

[1] 64 bytes at an SPI clock rate of 8MHz $\times$ 262.5Hz = 2.1 milliseconds per second not in sleep.

Table 9.3, and can be adjusted from 62.5Hz (64,000 'ticks' $\equiv$ 16msec) to 5kHz (800 'ticks' $\equiv$ 200μsec).

I discovered some problems with excessive noise when testing my prototype using the values provided as examples in the data sheet (revision F), the cause of which was an inadequate pause between ADC conversions. I fed this back to Texas Instruments who amended the sample timings in their next data sheet (revision G). For my final design I used the lowest sampling rate of 62.5Hz based on adjusting the default values in the data sheet to maintain a duty cycle of 25%.

Because the lowest sampling rate of the AFE4490 still greatly exceeds the design requirement of just 2 SPS (Section 5.5.3 on page 162), this allows for the possibility of applying digital filtering either on the AFE4490, which can average up to 16 samples per result according to a user-controlled setting, or in the ATMega328P itself.

9.8 Oxygen Saturation Sensor Circuit Design and Layout

The circuit design only required one revision during the prototyping stage to reduce the component count of the power supply. Rather than deriving power to the analog, digital and LED systems from three separate regulators, the digital subsystem and LEDs were changed to share a common higher output regulator.

Figure 9.18 below illustrates the circuit laid out across the top and middle layers in order to implement the oxygen saturation sensor, with the LED traces in blue and the photodiode lines in red. The photodiode lines are actively shielded on the bottom and top layers of the board by a guard ring driven by the AFE4490, shown here in orange:

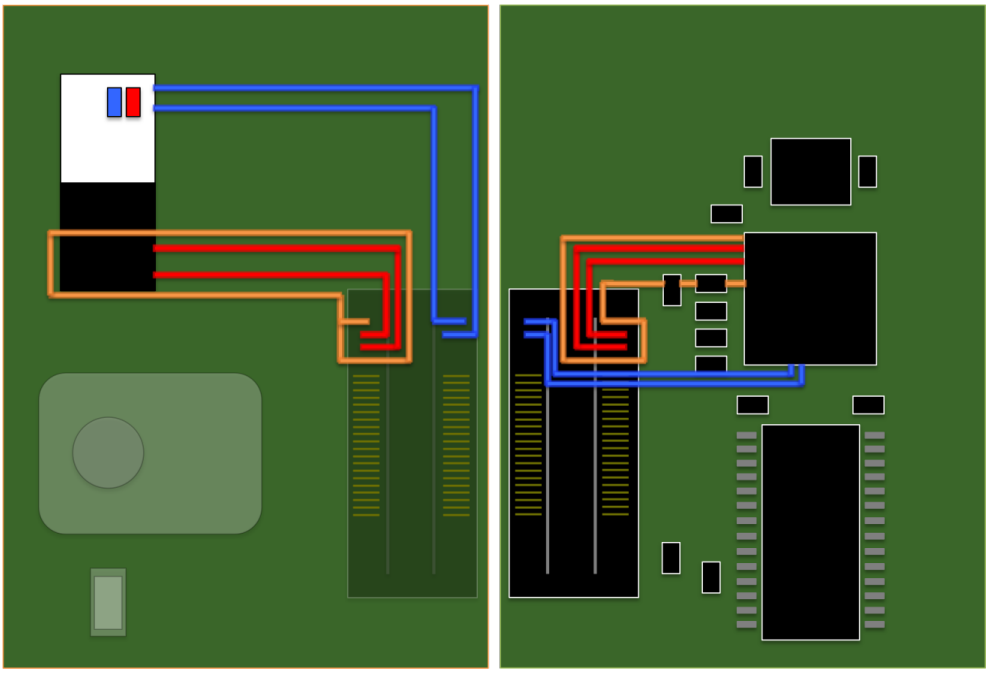


Figure 9.18: Layout of the connections between the DCM03 sensor on the top layer of the SmartPipe and the AFE4490 on the middle layer, with the LED connections in blue, the photodiode in red, and the guard ring in orange. Note the top layer is shown from the underside, so that the image 'opens' like a book down the centre

9.9 Testing the Oxygen Saturation Sensor

Initial Testing

The sensor was initially tested as a reflectance-based pulse oximetry sensor in order to develop the software and investigate initial TIA and LED settings. The results of one such test to measure the absorbance through my thumb is shown in Figure 9.20 using a >10mA photocurrent, 500kΩ TIA gain, 15pF feedback capacitor, and a sampling rate of 62.5Hz. It is clear that the system demonstrates good peak-to-trough range above the visible noise background in both wavelengths, with clear resolution of the heart rate, the upstroke of systolic ejection, and the dicrotic notch even at this lowest sampling rate.

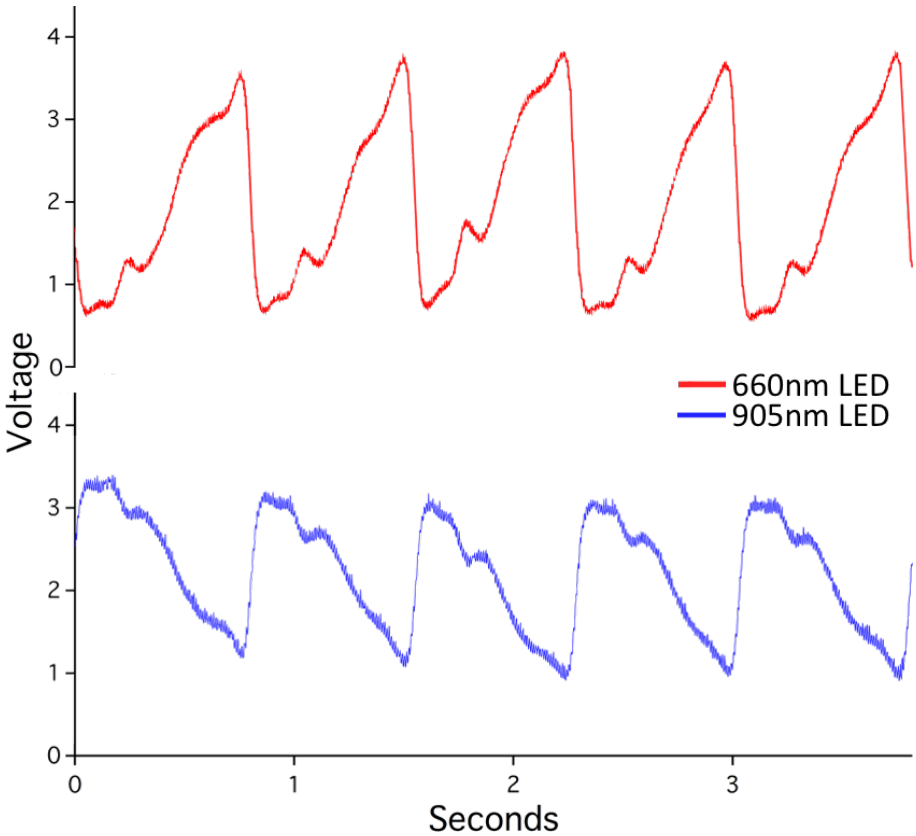


Figure 9.19: Testing the AFE4490 in a trans-cutaneous configuration. Note that the 660nm wavelength is most strongly absorbed by deoxyhæmoglobin, and 905nm by oxyhæmoglobin

Testing in the SmartPipe

Unfortunately, because the DCM03 sensor is in direct contact with the blood in the interface pipe, and the perfusion may not be pulsatile, the saturation sensor will not be able to internally normalise its results through the 'ratio-of-ratios' method employed by pulse oximetry. Furthermore, as discussed on page 276 in Chapter 8, simple dual-wavelength spectrophotometry is not a wholly accurate way of calculating the oxygen saturation of blood if one does not know the magnitude of scattering, the concentrations of other hæmoglobins present without co-oximetry (page 278), nor the original intensity. Likely sources of scattering in the SmartPipe include the blood itself, which will vary with the density of erythrocytes (hæmatocrit) and the inaccuracies in the 3D printing of the interface pipe which leaves microscopic ridges along the internal lumen of the pipe, thereby increasing scattering and reducing specular reflection.

The interface pipes were printed in a transparent material, and so in my initial experiments I painted the external surface black in order to prevent external light from reaching the photodiode. However, this appeared to completely abolish any response from the sensor. By instead painting the outer part of the pipe white prior to painting a few layers of

black, the photodiode was able to detect the internally reflected light to produce a signal. Presumably a polished mirror-like surface on the inner aspect of the flow tube would provide the optimal response.

I began by optimising the LED and gain settings so as to maximise the signal across the range of 0% – 100% saturation in defibrinated horse blood[1], which has the same absorption spectrum as human blood[140] and allow the subsequent development of a saturation calibration curve.

Manipulating Blood Oxygen Saturation

Initial attempts to manipulate blood saturation by exposing a volume of blood to gas mixtures in an enclosed volume were accompanied by some difficulties, including evaporation, precision control of the gas mixture, and rapid re-oxygenation when attempting to transfer the blood from the modified sample into the SmartPipe. A more sophisticated experiment involving a membrane gas exchanger in a closed-loop arrangement with a SmartPipe on the arterial and venous end would have been advantageous.

Nevertheless, I succeeded in employing an older technique of chemically deoxygenating the blood with sodium dithionite as reported in the historic literature, with the small caveat that this technique may generate small amounts of sulfhæmoglobin in the presence of H_2S[141] but that this was not expected to be present in my samples. Having measured the hæmatocrit of the samples at 45%, equivalent to a hæmoglobin concentration of 150g/L (2.33 mmol/L), I could fully desaturate each ml of blood with 9.3μmol of dithionite[2].

Figure 9.20: Sample vials containing oxygenated and chemically deoxygenated blood. Note the entrapment of air bubbles and drying of the blood on the wall which can potentially interfere with measurements

These experiments allowed me to produce Figure 9.21 containing the signal-to-noise ratios for multiple combinations of transimpedance and stage 2 gains versus the LED current.

[1] Sourced from TCS Biosciences Ltd., Botolph Claydon, Buckinghamshire, UK
[2] 4 O_2 molecules per hæmoglobin of molar mass 64,458g/mol

The signal level was the magnitude of the difference between saturations of 100% and 0% for each LED. From these I was able to determine that I could achieve an acceptable result by setting both TIA gains to 1MΩ and both LED currents to 3.13mA. The ×3 setting for the secondary gain stage of the 660nm LED began to saturate just beyond 3mA and so I chose a ×2 setting to remain within range. The 905nm LED provided its best result at ×3 secondary gain.

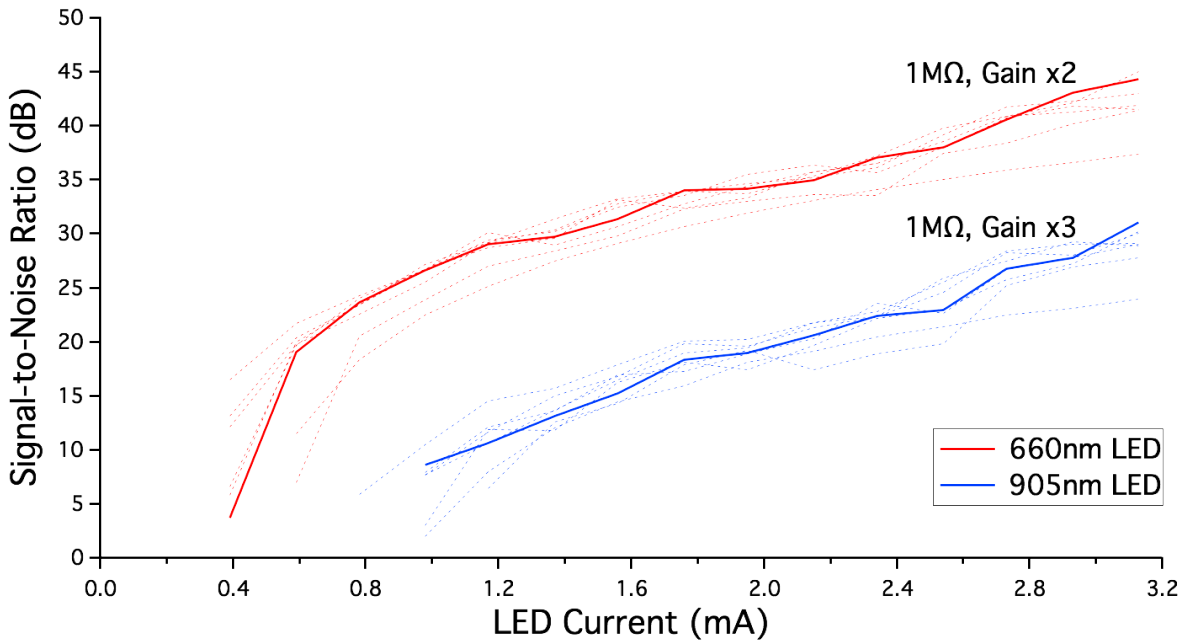


Figure 9.21: Signal-to-noise ratios of multiple LED tests in blood at 100% and 0% saturation with different TIA and secondary gain settings

A six-point calibration curve was then derived from the optimal LED values following volumetric mixing of the 0% and 100% saturated hæmoglobins to 50%, 70%, 80% and 90%, shown below in Figure 9.22, with some of the noise undoubtedly arising from the DAC-PowerLab system (page 190)[1].

[1]The standard deviations are taken from the first order Taylor expansion of the ratio of standard deviations for two independent variables: $|\text{SD}| = \text{Ratio}_{\frac{660}{905}} \cdot \sqrt{\left(\frac{\text{SD}}{\text{Result}}\right)^2_{660nm} + \left(\frac{\text{SD}}{\text{Result}}\right)^2_{905nm}}$

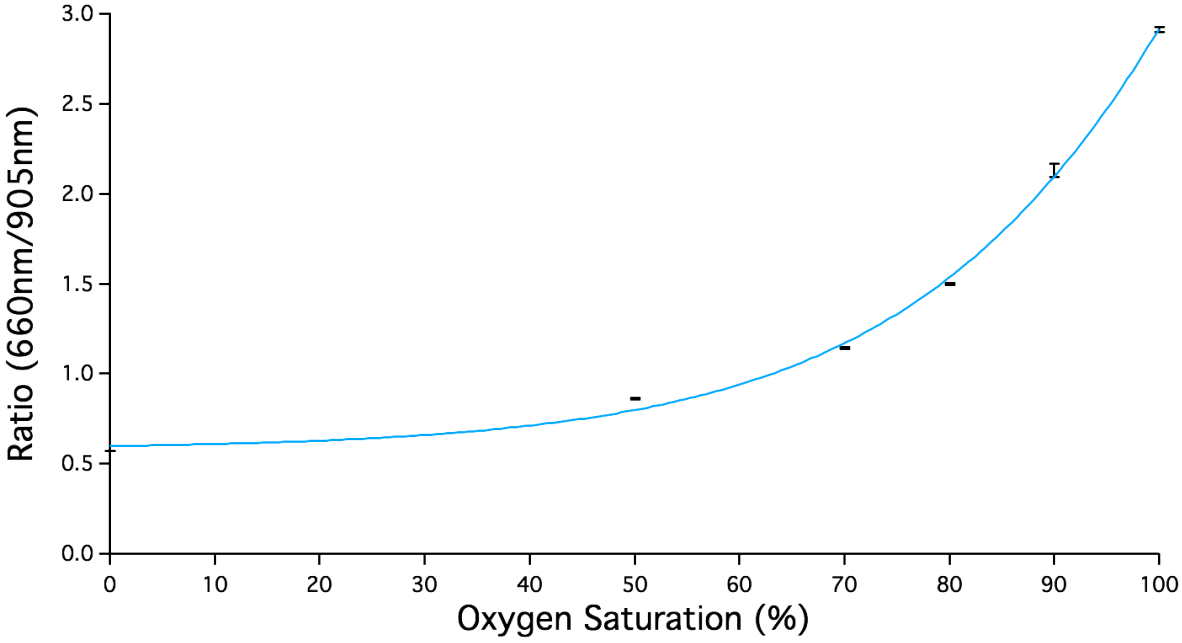


Figure 9.22: Six-point S_pO_2 calibration curve derived by mixing quantities of 100% saturated and chemically-desaturated horse blood, showing that the system is most sensitive in the biologically important range from 70% – 100%

Developing this initial work into a more complete calibration curve with 1% accuracy from 75% to 99.9% would take a more sophisticated apparatus with and the ability to accurately mix volumes of various hæmoglobins without exposure to ambient air or excess dithionite.

Despite these drawbacks, I was able to successfully develop and test the oxygen saturation sensor to the extent that it could determine the extremes of the saturation scale with good full-scale resolution and noise performance, using a miniature fully-integrated blood-contacting reflectometric sensor. The system bandwidth was also far in excess of that required by the initial design considerations in Chapter 5, allowing for further improvements to the noise performance by software filtering and cross-correlation with the pressure sensor for gating if used in a pulsatile system.

9.10 Software Power Savings for the LMP90080 and AFE4490

I was able to realise improved power savings by taking advantage of the wide degree of customisation possible within the software of the LMP90080 and AFE4490. One unavoidable current draw was the use of the RTD and pressure sensor buffers on the LMP90080, requiring a total of 700μA of extra current. Interestingly, adjusting the sampling rate did not appear to have any effect on the total power requirement.

Minimising Sensor Currents

The first change was to reduce the magnitudes of the current for driving the RTD and pressure sensors to 500μA, discussed previously on page 291, saving up to 1mA in total draw. This has the benefit of also reducing any RTD error due to self-heating (page 230 in Chapter 8), but this may not compensate for the decreased signal magnitude.

The current draw of the DCM03 was also greatly reduced from the initial use 10mA following the experiments that produced Figure 9.21 on page 314.

Adjusting LED Timings

The data sheet timings for the AFE4490 provide a LED flicker with a 25% duty cycle, even at the lowest sampling rate of 62.5Hz. I adjusted these to instead reflect the 1,600 'tick' `ON` duration required for a 2.5kHz sampling rate, whilst retaining the total sampling period of each of the 4 blocks the equivalent of 62.5Hz, as shown in Figure 9.23 below. This created a signal with a duty cycle of 2.5%, which could have been pushed further to 1.25% if required by using the minimum 800 'tick' duration equivalent to a 5kHz rate. This single change reduced the current drawn by the entire system by 5mA.

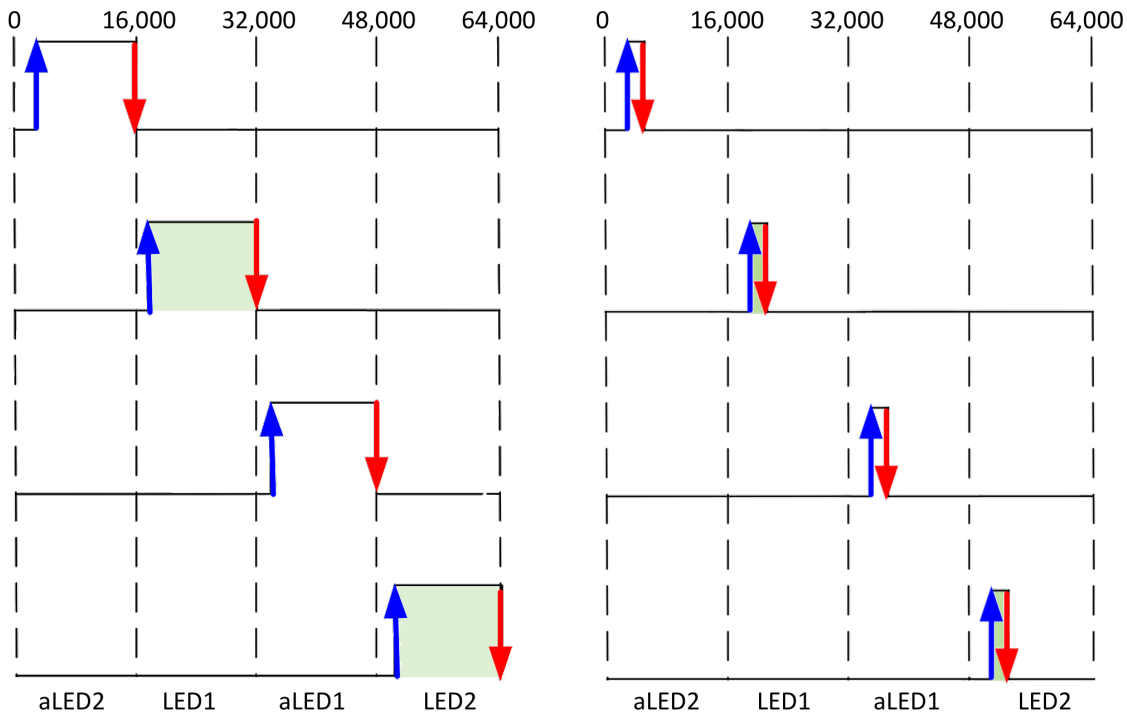


Figure 9.23: Left: Original 62.5Hz timings with 25% duty cycle **Right:** Modified 62.5Hz timings with 2.5% duty cycle. Note the dramatic reduction in the active time of each LED highlighted in green

Look-Up Tables

As discussed previously on page 308, the longer that the ATMega328P is able to spend in a power saving sleep mode, the less current the system will draw. One particular demand placed on this limited 8-bit microcontroller is the derivation of the temperature from the voltage across the RTD using the Callendar-Van Dusen equation (page 227) at least 26 times per second, involving both floating-point values and square roots. This is even before filtering the results, requiring even more floating-point manipulation.

Where a variable can only take on a limited range of potential values, or a sensor has a constrained response, it is possible to pre-calculate each possible result in advance and place these values into system memory as a look-up table instead of undertaking time-consuming live calculations. Upon receipt of a new reading from the sensor, the system then scans through the LUT to find the nearest matching result, and returns this to the user.

I therefore benchmarked the time required for calculating the RTD's temperature by the quadratic formula versus that required for using a look-up table. I implemented two different search algorithms for the LUT - the first described above as a simple marching comparison from the first value onwards, and a second using a binary dissection tree which should reduce the total number of fetch-and-compare operations. These results are given below in Table 9.4.

The final LUT contained 451 values derived according to the system gain, reference resistor value, and the characteristics of a Class A 3850ppm/K RTD – one for each of the expected LMP90080 ADC output values from 0°C to 45°C, in 0.1°C steps.

Test	Time (μsec)	Clock Cycles
Solving the Quadratic Formula	286.6	2290
Binary Dissection in FLASH	25.3	202
Binary Dissection in RAM	18.2	145
Linear Search in RAM	0.86	7

Table 9.4: Average times taken to directly calculate the temperature using the quadratic formula versus using a look-up table with various algorithms, from 1,000 tests. For reference, solving a logarithm takes 2.3μsec, or 18 clock cycles

Summary of Power Savings in Software

After applying all of the changes listed in this section, the entire SmartPipe system drew 5.7mA for a total lifetime of ≈17 hours on a single 110mAh battery. This could be improved further by applying some of the changes recommended below.

9.11 Conclusions and Future Work

This project has made significant progress towards creating a fully-integrated, multi-modal physiological sensor suite on the scale of a through-flow pipe for a blood perfusion circuit. The temperature and pressure sensors function according to the design specification, and the oxygen saturation sensor demonstrates the pre-requisite dynamic range and low noise required to produce a sensitive calibration curve in order to finalise the design.

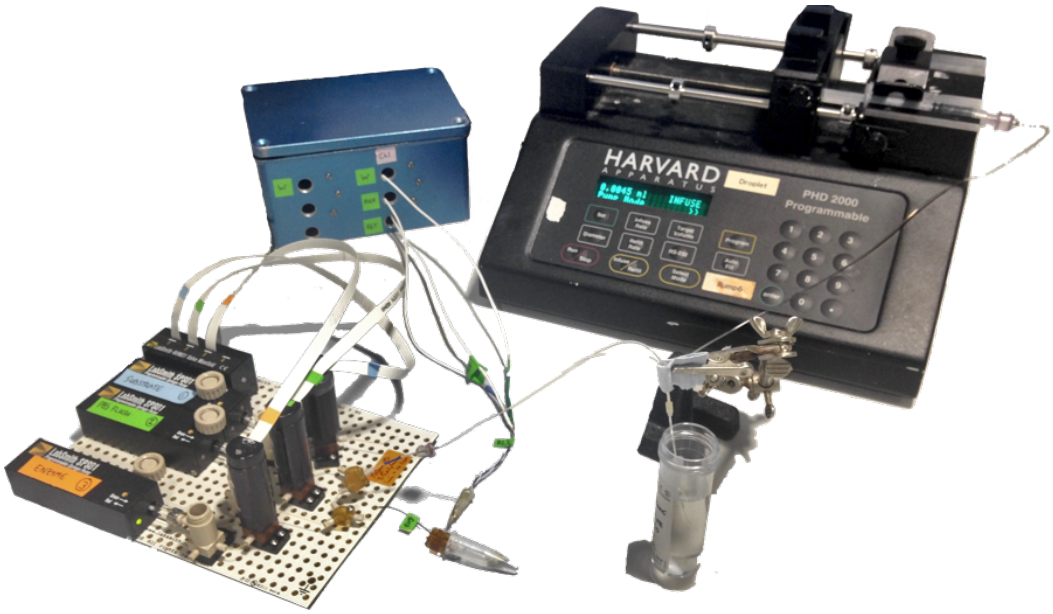
Future work in this area should return to the flow sensor, and investigate the digital system with a unified ambient/sensor/actuator design that I proposed on page 251 in Chapter 8. Alternatively, if it became possible to source a waterproof MEMS ultrasound transceiver and matching layer, the current state of integrated circuitry for RF detection should make it possible to integrate a doppler flow sensor into a future version of the SmartPipe.

With respect to the oxygen saturation sensor, since I commenced the SmartPipe project, APM Korea has now released a quad-wavelength DCM06 device (570nm, 660nm, 810nm and 905nm), and an ultra-miniature version of the DCM03 called the DCM05 (4.8mm$\times$2.6mm$\times$1mm), extending the possibility of performing basic co-oximetry, monitoring the hæmatocrit (at 810nm), or further reducing the size of the system. Texas Instruments has also released an even smaller, lower power AFE4404 device in a 2.6mm$\times$1.6mm$\times$0.5mm 15-ball BGA package with an integrated 4MHz oscillator for personal healthcare devices. The higher TIA gain (up to 2MΩ) and dynamic power savings requiring just 650μA would reduce the current draw even further, pushing up the battery life of the SmartPipe.

If this were combined with a faster microcontroller, perhaps with an integrated wireless controller, the system could switch from being interrupt-based to on-demand real-time clock based wherein the microcontroller demands single-shot conversions from its peripherals on a regular basis and deactivates them in between. The higher speed would mean that SPI and I^2C queries, as well as data conversions, could be completed rapidly and maintain the system in a sleep state for a greater proportion of the time.

Overall there are a many exciting avenues that remain open for future development of this first version of the SmartPipe in order to achieve the goal of improving and commoditising normothermic organ perfusion.

Part 3:
The Live Renal Function Monitor



Chapter 10

Developing a Real-Time Renal Function Sensor

This Chapter introduces the concepts underlying the monitoring of renal function and the background of existing sensing techniques.

The following Chapter will present the tools and techniques I developed myself during this Thesis, and the results from a real-time creatinine sensing system, with a discussion of the potential future direction of research in this area.

10.1 Design Requirements

Section 2.7 in Chapter 2 of the Introduction (page 66) introduced the concept of 'renal function' and described what it means in clinical practice in relation to single measurements of the serum creatinine concentration.

Within the kidney, the nephron comprises a complex array of specialised anatomical subdivisions, each concerned with particular functions to regulate the *'milieu interieur'* of the human body. In time we will undoubtedly see assays developed which aim to identify specific renal pathologies from combinations of the wide array of potential biomolecules listed in Figure 2.7 on page 67, or use other molecules that are solely filtered at the glomerulus (such as inulin), but I felt it still worthwhile to focus on replicating the standard clinical practice of measuring creatinine as a marker for global renal filtration function, yet adapting it to a real-time system to monitor for momentary deterioration in an isolated normothermic perfusion system.

The reasons include (i) a broad base of literature supporting creatinine as a useful marker of renal function, (ii) its use in worldwide clinical practice leading to immediate familiarity with the results of such a monitoring system for any clinician, (iii) the suitability of the underlying chemistry of the detection methodology and (iv) the challenge of adapting this into a real-time system.

The overall design concept was to create a portable, low-cost, largely turn-key, miniature system for continuously sampling and assaying normal creatinine concentrations in either the blood or urine of an isolated perfused kidney. This requires a system capable of detecting concentrations between 60μm – 120μm for blood and 7– 16mM in the urine (Table 2.1 on page 52 in Chapter 2). Note that these blood creatinine concentrations are only $^{1}/_{25}{}^{\text{th}}$ – $^{1}/_{50}{}^{\text{th}}$ the concentration of blood glucose, and that there are no systems presently capable of continuous real-time creatinine monitoring in a clinical setting[142].

The rest of this Chapter will discuss the methods used for measuring creatinine in clinical practice, the challenges of adapting these methods to a continuous real-time system, and the experience within our laboratory for creating such systems. In the next chapter I will highlight where I have developed innovations and modifications to existing techniques, and present the results of testing the system.

10.2 Measuring Creatinine in Clinical Practice

Creatinine is the second most common analyte in clinical chemistry after glucose[143]. Three main techniques are presently used for measuring creatinine concentrations in

clinical samples: (i) the Jaffe reaction, (ii) enzymatic methods, and (iii) isotope dilution mass spectrometry. We will begin with the latter, against which all other methods are now compared.

Isotope Dilution Mass Spectrometry (IDMS)

This is considered the most accurate method of quantifying analytes of interest in modern clinical biochemistry. The principle is simple, and akin to estimating wild populations of animals by tag-and-release methods. Beginning with a sample of unknown quantity yet known isotopic composition and diluting it with a standard of known quantity and isotopic composition, one is able to determine the concentration in the original sample by measuring the final dilution ratio of the isotope in question. This method combines internal ratiometric normalisation with the high precision and low limits of detection of modern mass spectroscopy, leading to highly accurate and reproducible results with low bias.

Following the 2006 publication of a report[144] by the US Laboratory Working Group of the National Kidney Disease Education Program in collaboration with the European Communities Confederation of Clinical Chemistry[1], IDMS[2] is now considered the reference method for creatinine measurement. The 1998 EU Directive 98/79/EC requires all laboratory calibration to be traceable to a higher standard, and similar laws have been enacted in the US since 2011. As a result, all manufacturers of clinical test equipment for creatinine must use calibration methods that can be traced back to an IDMS standard.

The MDRD formula for estimating the glomerular filtration rate (GFR) from creatinine measurements (page 72 in Chapter 2) now reflects this standardisation[74], which improves the differentiation between earlier stages of chronic kidney disease (CKD 1–3).

Unfortunately, the methodology and the size and cost of GC-MS devices required to undertake this assay did not lend themselves to miniaturisation for incorporation into a continuous in-line creatinine assay for isolated perfused kidneys. I therefore explored the possibility of using one of the other two methods in my system.

[1]Now the European Federation of Clinical Chemistry and Laboratory Medicine
[2]Often Isotope-dilution gas chromatography mass spectrometry (GC-IDMS)

The Jaffe Reaction

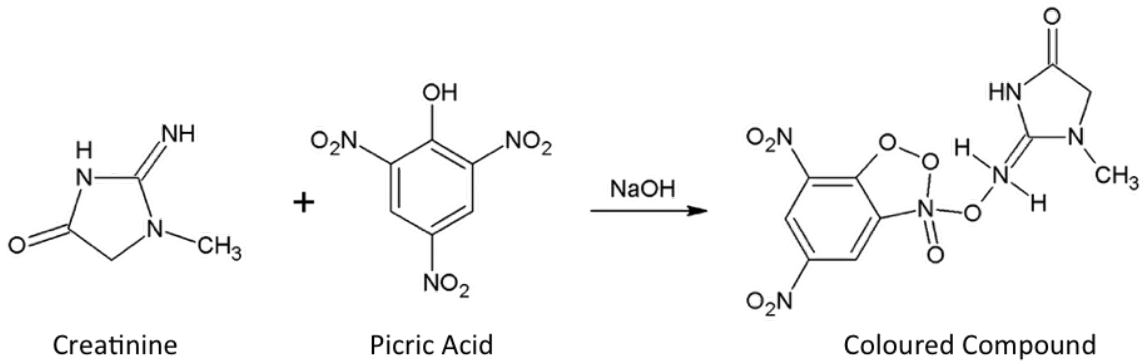


Figure 10.1: The Jaffe reaction between creatinine and picric acid to produce a coloured compound under alkaline conditions

This method of detecting creatinine predates its recognition as an important indicator of renal function. The first formal description of the production of a coloured compound following the alkalinisation of the reaction product of picric acid and urinary creatinine was published by Max Jaffe (1841 – 1911) in 1886, for whom this detection method is named. The intensity of the colour change, and thus the amount of creatinine in the sample, could be rapidly assessed with colourimetry or spectrophotometry with maximum absorbance at 520nm. The reaction is shown in Figure 10.1 above.

This reaction was first used as a method for quantifying the creatinine in human urine samples by Otto Folin (1867 – 1934) as early as 1902[145]. Folin was one of the earliest pioneers of clinical biochemistry, and his interests lay in studying the principles of human protein metabolism. He reasoned that quantifying the nitrogenous byproducts within the urine of adults fed specialised diets could provide insights into human disease. Using the Jaffe reaction and a Duboscq colourimeter, Folin was the first to prove that there is a basal production of creatinine into the urine which he termed the 'endogenous metabolism', and was quite separate from the production of urea which was more dependant on protein consumption.

In the 1920s, Poul Rehberg (1895 – 1989) correctly surmised that this basal excretion rate of creatinine could be used to quantify underlying renal function[69], forever cementing the success of Folin's work and leading to the continued use of the Jaffe reaction for the laboratory assessment of creatinine in blood and urine samples to this day. The reasons for its persistence, particularly in the developing world, include its low cost and simplicity, requiring minimal sample preparation yet producing a result that is readily amenable to electronic photodetection or manual colourimetry for rapid quantitation.

Unfortunately, the method is not without its drawbacks, not least of which is the historic lack of standardisation between laboratories, demonstrated by the work leading to the

development of the IDMS standards[144]. The Jaffe reaction is also highly non-specific, and can produce false positives or negatives with a vast number of endogenous and exogenous compounds often found in human samples, including trace amounts of protein, glucose, ketone bodies, bilirubin and certain aminoglycoside and cefalosporin antibiotics.

Attempting to calibrate against these can in fact introduce greater uncertainty to values on the borderline between normal and abnormal function, and paradoxically underestimate the creatinine concentration of urine where none of the endogenous interferents are found[146]. As an illustration of the impact of even small errors, an increase of just 20μmol in the absolute value of the serum creatinine concentration can mean the difference between normal function and early renal failure.

For these reasons, the Jaffe method is being slowly replaced by enzymatic detection in the developed world, explaining why I did not try to pursue this further.

10.2.1 Enzymatic Methods

There are two main methods for the enzymatic detection of creatinine in use today, each employing a different enzyme reaction scheme yet relying on the same principle of substrate specificity in order to generate a product that is detectable with low rates of interference (noise), resulting in a system with high accuracy, reproducibility and low limits of detection. In the case of creatinine sensing, this specificity reduces the effects of many substances that would typically interfere with the Jaffe reaction, including albumin, glucose, bilirubin, haemoglobin, pyruvate, ketones, and catecholamines (particularly dopamine or dobutamine which are used in an intensive care setting). Indeed, recent publications have indicated that enzymatic methods approach the precision and accuracy of IDMS[146, 147].

High grade purified forms of the enzymes used in both of the following systems are available from commercial suppliers[1,2].

Method 1: Creatinine Deiminase

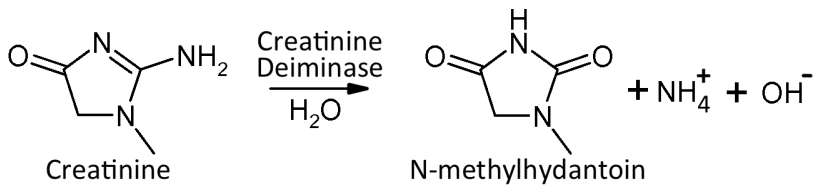


Figure 10.2: Creatinine deiminase catalyses the production of N-methylhydantoin and ammonium from creatinine

[1]Toyobo Co., Ltd., Osaka, Japan
[2]Sigma-Aldrich UK, Gillingham, Dorset, UK

This reaction, shown in Figure 10.2 above, was first pioneered in 1983 by Guilbault and Coulet[148]. It relies on the creatinine deiminase enzyme derived from *Clostridium spp.* to catalyse the single-step conversion of creatinine into ammonium and N-methylhydantoin in a 1:1:1 molar ratio.

Method 2: Three-Enzyme Digestion

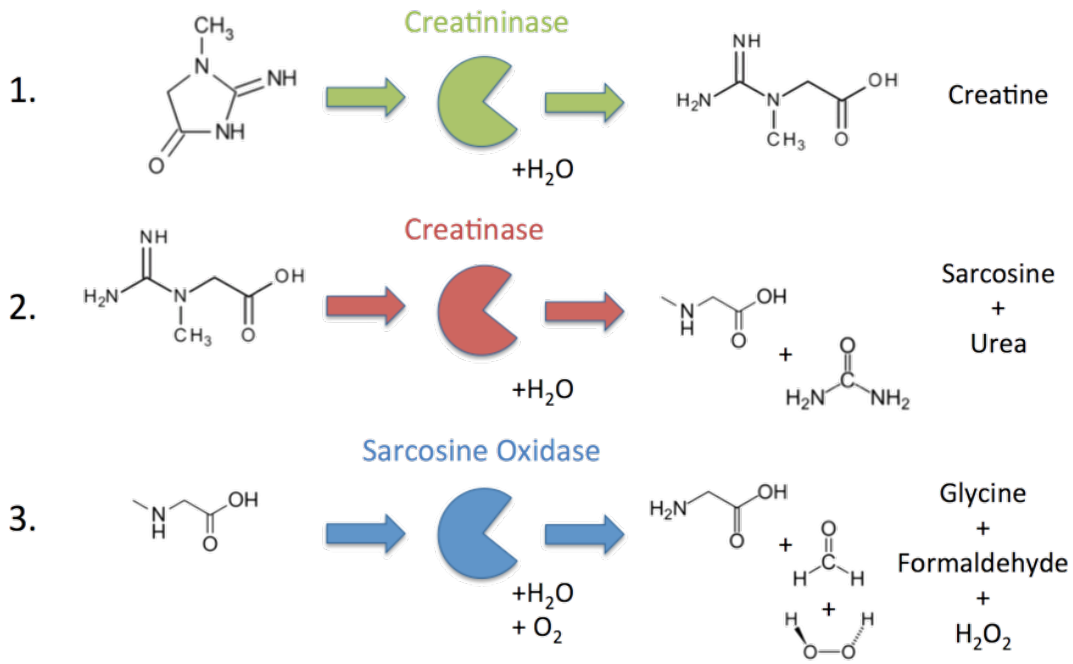


Figure 10.3: The three-enzyme digestion of creatinine into glycine, formaldehyde and hydrogen peroxide, with urea as an intermediate byproduce

This method, shown in Figure 10.3 above, was first proposed by Tsuchida and Yoda, and also published in 1983[149]. It relies on a more complicated three-step digestion of creatinine into hydrogen peroxide, formaldehyde and glycine in a 1:1:1 molar ratio, via creatine and sarcosine as intermediates, and urea as an intermediate byproduct.

10.3 Selecting an Enzymatic Method for the Real-Time Assay

Our research group has many years of experience in developing microfluidic real-time enzymatic assays of tissue glucose and lactate levels for monitoring cellular energy balance[150–152], which gave me confidence that an enzymatic method could be translated into a real-time system for monitoring renal function in isolated perfused kidneys.

We have learned through experience is that it is important to consider the detection method for the reaction product at an early stage in the design process. Both of the reaction schemes for creatinine deiminase and the more complicated 3-step process of Tsuchida and Yoda produce species that are amenable to either electrochemical or spectrophotometric quantitation. Of these two, electrochemical methods are more suited to miniaturisation owing to the problems of optical path-lengths at small scales and the creation and stability of monochromatic light sources required for colourimetric or absorption-based detection.

I considered the possible electrochemical detection methods for each reaction product in turn, keeping in mind the design requirements of a turn-key, miniaturisable, real-time system.

Monitoring the Creatinine Deiminase Reaction

There are four possible non-optical methods for monitoring the ammonia produced by this enzyme. The first two of these rely on further enzyme-coupled reactions, and the second two rely on directly detecting the ammonia resulting from the catalytic breakdown of creatinine.

Coupling to Glutamate Dehydrogenase and Detecting NAD(P)H

The first scheme couples the first stage catalysis to a glutamate dehydrogenase (GDH) enzyme which incorporates ammonium into glutamate and oxidising a cofactor in the process in a 1:1:1 molar ratio. The cofactor for GDH is either NADH or NADPH depending on the specific enzyme selected.

$$\text{2-Oxoglutarate} + \text{NH}_4{}^+ + \text{NAD(P)H} \xrightarrow{\text{GDH}} \text{L}-\text{Glutamate} + \text{H}_2\text{O} + \text{NAD(P)}^+$$

The first problem with detecting ammonia from biological samples is the need for a pre-treatment stage to remove any endogenous ammonia[153]. This may not be miniaturisable or suitable for real-time sensing.

The second concern with this methodology is the complication of measuring NAD(P)H. Existing electrochemical methods generally rely on non-convential materials, such as boron doped diamond, glassy carbon or carbon nanotubes functionalised with gold nanoparticles or redox mediators such as Prussian Blue or Meldola's Blue. Each of these methods has problems in terms of manufacturing costs, poor selectivity and short lifetimes in real biological samples[154].

Coupling the Second Reaction to Glutamate Oxidase

In this system, the L-Glutamate produced by the glutamate dehydrogenase (GDH) is coupled into a further oxidation reaction by glutamate oxidase (GLOD) as follows[155]:

$$\text{Creatinine} + \text{H}_2\text{O} \xrightarrow{\text{Deiminase}} \text{N}-\text{methylhydantoin} + \text{NH}_3$$

$$\text{2-Oxoglutarate} + \text{NH}_4{}^+ + \text{NAD(P)H} \xrightarrow{\text{GDH}} \text{L}-\text{Glutamate} + \text{H}_2\text{O} + \text{NAD(P)}^+$$

$$\text{L}-\text{Glutamate} + \text{H}_2\text{O} + \text{O}_2 \xrightarrow{\text{GLOD}} \text{NH}_3 + \text{2-Oxoglutarate} + \text{H}_2\text{O}_2$$

Note that this effectively reverses the intermediate GDH step, regenerating all of the reaction products without amplification, and consumes oxygen in the process. This consumption can be monitored with a Clark-type oxygen electrode, as introduced on page 273 in Chapter 8, or amperometrically through the reduction of H_2O_2.

Unfortunately, whilst the nature of O_2 and H_2O_2 sensing is well understood, this method still needs a pre-treatment step to exclude endogenous $NH_4{}^+$ from interfering with the reaction, and careful attention to the pH of the system to control the ionisation of $NH_3/NH_4{}^+$.

This method also lacks any real benefits over the three-enzyme system of Tsuchida and Yoda[149] in terms of complexity or predicted sensitivity. For an ideal enzymatic sensor, the equilibrium of the system should be far to the right, yet because of the circular nature of the reaction between the second and third steps, the system may be rate limited by excessive NH_3 over time, or similarly by the consumption of NAD(P)H in the production of L-Glutamate.

Finally, this method was excluded from a practical standpoint because of the excessive cost of the glutamate oxidase enzyme at £200–£300 for just 5 units[1].

Potentiometric $NH_4{}^+$ Sensing

An alternative method for monitoring the deiminase reaction is to measure the ammonium ion in solution with a potentiometric sensor. Unfortunately, the selectivity of such sensors is poor owing to the similar diameter of the $NH_4{}^+$ cation (296pm) and the more common K^+ cation (282pm). As a result, the most common natural ammonium ionophore, Nonactin, only has a log selectivity of $NH_4{}^+$ over K^+ (log $K^{POT}_{NH_4{}^+,\ K^+}$) of -0.8. This indicates that a change in the K^+ concentration of 6.3-fold will give the same voltage response as a 1-fold

[1]Pricing from Sigma-Aldrich UK, Gillingham, Dorset, UK

change in the NH_4^+ concentration. Even the most selective synthetic ammonium ionophore only exhibits 10× greater selectivity over K^+[156].

Referring back to Table 2.1 on page 52 in Chapter 2, we can see that that the ratio of urinary K^+ to NH_4^+ can vary over an 18-fold range from 1:3 to 6:1, and that of serum K^+ to NH_4^+ from some 30:1 to 80:1. Unless the system incorporates an ion exchange mechanism to greatly suppress the endogenous K^+ concentration[157], or a simultaneous K^+ sensor to provide a compensation mechanism, it would be impossible to measure the creatinine by this method in a perfusion system. Furthermore, relying on two sensors operating simultaneously to detect one analyte merely compounds errors of stability and sensitivity over time. For these reasons this methodology was excluded at an early stage.

Detecting NH_3

The final method of monitoring the creatinine deiminase reaction is to ensure the solution remains basic and to detect any ammonia (NH_3) either in gaseous form or in solution. There are two potential sources of difficulty with ammonia sensing which led me to exclude this method, and hence the use of the creatinine deiminase system entirely.

The first of these is the need to design a miniature ammonia sensor *de novo*. There are a diverse range of non-optical methods described in the literature, including amperometry with electroconducting polymer films[158], and rugged solid-state metal-oxide sensors. However, many of these methods appear to have problems with selectivity, limited accuracy, and loss of sensitivity over time as the reactive surface becomes poisoned by the analyte[159].

My second concern was the pH and temperature stability of the reaction system. Examining the supplier's data sheet for creatinine deiminase[1] shows that the enzyme functions optimally around pH 9, with the activity falling to 75% of the peak above pH 9.25, the pK_a of NH_3/NH_4^+. At pH 9.0, ≈36% of the ammonia will be in the form of NH_4^+ and unavailable to the sensor, but a change of 0.1 pH in either direction can shift the freeNH_3 concentration over a range of 10%.

Furthermore, the enzyme functions optimally at a temperature of 65°C – 75°C, with its response falling to <30% at temperatures below 30°C, perhaps driven by the evolution of the product (NH_3) from solution at these higher temperatures. This is perhaps more troublesome than creating a strongly buffered basic solution for the reaction system, because there are no simple ways of creating a thermostatting system on the miniature scale required for a continuous real-time creatinine assay, and this would have taken the bulk of my work in this area away from actually developing the chemistry of the system.

[1] Toyobo Enzymes, number CNI-311

Monitoring the Three-Stage Digestion Reaction

The system proposed by Tsuchida and Yoda results in two potential targets for non-optical detection – urea and H_2O_2.

Detecting Urea

The first of these, urea, requires a further coupling reaction to urease which catalyses the production of NH_3 and CO_2 as shown below:

$$\text{Urea} + \text{H}_2\text{O} \xrightarrow{\text{Urease}} 2\,\text{NH}_3 + \text{CO}_2$$

Whilst the detection of NH_3 is complicated by difficulties previously discussed in Section 10.3 on page 328, CO_2 production can be quantified with a more common Severinghaus electrode, or more exotic doped nanomaterials[160].

The Severinghaus electrode requires a particular internal configuration of a glass pH electrode encased within a solution of $NaHCO_3$ of known pH, and separated from the sample solution by a gas-permeable membrane. As CO_2 passes through the membrane, it dissolves into the $NaHCO_3$ solution to evolve H^+ ions. These are then potentiometrically sensed at the internal pH electrode. This is shown in Figure 10.4 below.

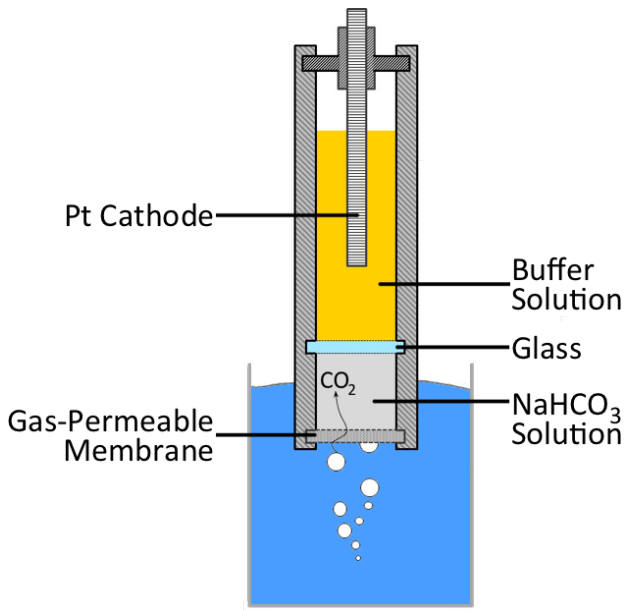


Figure 10.4: Image of a Severinghaus-type electrode demonstrating the platinum cathode, the internal $NaHCO_3$ filling solution and the glass membrane. The reference electrode is not shown in this figure.

Whilst the principle is well understood, and the sensor behaves in a linear manner over the normal range of human blood pCO_2, this triple-walled, liquid-containing sensor is very hard to miniaturise, and only a handful of reports exists in the literature of micromachined Severinghaus-type electrodes, with very slow response times[161].

Furthermore, the amount of CO_2 evolved from the complete digestion of creatinine will be in the sub-millimolar range at best. This means that any CO_2 sensor system will be exposed to an offset of tens of times the expected signal magnitude from the background levels of CO_2 dissolved in the sample from normal metabolism (4.5 – 6 kPa, $\equiv$ 1.75 – 2.33 mmol), whether that sample is drawn from the blood or the urine.

Finally, any biological sample will also contain levels of urea that are also far greater than that of creatinine (see Table 2.1 on page 52 in Chapter 2 for further details). Reports do exist of interesting simultaneous assay systems relying on post-processing of the data arising from a single sensor fed by a combination of reaction beds and delay lines[155], but these are beset by problems of reproducibility, long setup times and problems with flow dependence.

Detecting H_2O_2

This reaction product is the most amenable to electrochemical detection. It does occur within the blood and urine as a result of oxidative metabolic processes, but at low micromolar levels which rapidly diminish under the effects of endogenous antioxidants in the plasma including catalase, hæme, and ascorbate[162]. Thus, the only appreciable source of H_2O_2 in the triple-enzyme scheme is the creatinine itself via sarcosine oxidase.

H_2O_2 is also readily detectable through amperometry, which is where much of the experience in our laboratory lies, particularly with respect to the creation of stable, reproducible and ultra-miniature amperometric sensors.

Finally, the overall equilibrium of the triple-enzyme system lies far to the right with the generation of products and consumption of the substrate, unlike that of detecting creatinine deiminase via glutamate dehydrogenase and glutamate oxidase. This indicates that a higher potential level of product, and thus signal, will result from a smaller quantity of substrate, improving the signal to noise ratio and limits of detection for this system.

For these reasons, I chose to implement the three-enzyme process of Tsuchida and Yoda for the live monitoring of renal function, through the amperometric detection of H_2O_2.

10.4 Developing the Real-Time Assay System

The glucose and lactate sampling systems developed within our laboratory leverage a combination of microdialysis, microfluidics and amperometric sensing to create robust continuous-flow real-time assay systems. I chose to apply this combination to the three-enzyme process of Tsuchida and Yoda.

Microdialysis

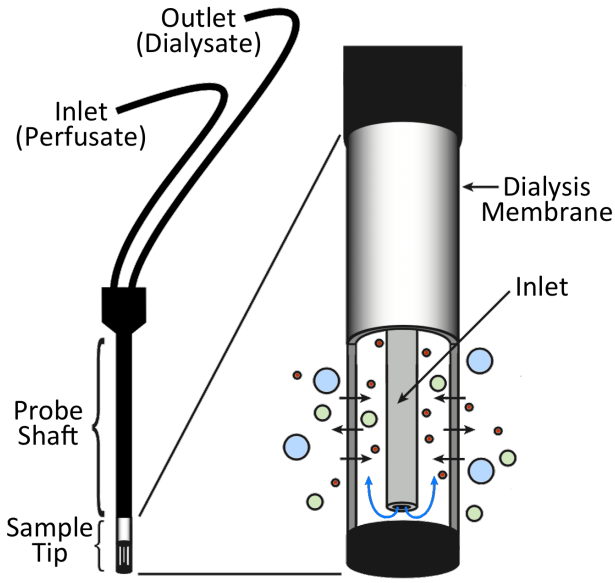


Figure 10.5: Illustration of a linear microdialysis probe demonstrating the principles of microdialysis. The perfusate enters the probe through a central channel made of impermeable material before washing back over the internal surface of the semi-permeable membrane and into the probe's return line for collection and analysis. Adapted from [163]

Microdialysis is a method for obtaining continuous samples of small molecules from a tissue or solution of interest, whilst minimising interferents, and was originally pioneered in the 1970s for sampling neurotransmitters from the rat brain[164]. It works by continuously perfusing one side of a semi-permeable membrane with a fluid which lacks the molecule(s) of interest so that target molecules will diffuse down their concentration gradients across the membrane into the *perfusate*. At the same time, molecules above the cut-off weight of the membrane, or which are already in equilibrium with the perfusate, will not change in concentration. The post-membrane *dialysate* then carries the target molecule to the detection system. An example of a linear microdialysis probe is shown in Figure 10.5 above.

The benefits of microdialysis arise largely from the molecular-weight selectivity of the semi-permeable membrane, allowing the exclusion of larger interferents such as cells, lipids and proteins from the sampling stream, and the ability to adjust the perfusate to 'capture' specific molecules of interest. This exclusion is particularly important when sampling from contaminant-rich biological liquids such as blood or urine and is the reason for selecting this method of sampling for the final envisaged system.

The ability of the probe to recover the analyte of interest is influenced by a number of factors including its total surface area, the thickness and cut-off value of the membrane, the flow-rate of the perfusate, the temperature and pH of the tissue and perfusate, and the pK_a of the analyte, which influences its ionisation[163]. Often, the 'recovery'[1] will be less than 100%, yet it is possible to perform calibrations *in vitro* to compensate for the degree of recovery if necessary.

In static cell-dense environments such as the brain, these simple *in vitro* calibrations may not correspond to the working environment of the sensor owing to the long diffusion path-lengths, low effective sampling volumes, and consumption or active uptake of molecules of interest, requiring further methods such as *no net flux* (NNF) to derive the true recovery. However, in well-mixed liquid samples such as the blood or urine, these exclusionary barriers do not apply and the situation is much closer to that of the homogeneous *in vitro* calibration solution. In addition, in an isolated perfused kidney preparation with minimal to no endogenous production of creatinine through protein metabolism, it is possible to imagine a standard addition experiment within the closed system to measure the recovery if necessary.

Microfluidics

The term 'microfluidics' describes the practice of working with volumes of liquid at or below the nanolitre scale, with flow channels only tens to hundreds of microns in diameter. Unlike traditional laboratory analyses, operating on these scales brings powerful advantages in terms of reducing the required volumes of samples and potentially costly reagents, whilst improving sensitivity, reproducibility and the speed of analysis[165]. This is particularly useful for enzyme-based reactions, where the enzymes themselves may be particularly costly, and where only small amounts of substrate may be available, as is the case with microdialysis.

Microfluidic devices do have some drawbacks in terms of the scale of the precision pumps and dosing systems required for manipulating the substrate and reagents, which can be many times larger than the reaction or detection chamber itself. There are also interesting

[1] The ratio of the true concentration within the sample to the concentration in the dialysate

problems with translating traditional macro-scale systems down to these volumes, where the viscous forces of water dominate over the inertial, leading to fully laminar flows with poor mixing which only occurrs through diffusion and not convection[165].

Droplet Microfluidics

Where the microfluidic system incorporates long delay lines to ensure full mixing, or simply to bridge the distance between the sampling probe and the detection system, viscosity will lead to Taylor dispersion, whereby initially focal step changes in concentration can be smeared out along the length of the tubing, decreasing the magnitude of the signal at the detector[166]. This results from the non-uniform cross-sectional velocity profile of a bolus as it passes along the tube, with wall effects slowing the centre of the bolus through friction. One innovation to overcome this effect is to use droplet-based, 'digital' microfluidics wherein aliquots of reagent and substrate are mixed and then broken up by a hydrocarbon or fluorocarbon oil during transportation. This system does lead to better internal mixing of the contents of the droplets in flow, yet requires more sophisticated 3D microstructures to strip the droplets of oil to allow electrochemical sensors to function[166]. This particular form of microfluidics was not used during this project, but could be incorporated into a future system requiring physical separation of the sampling probe from the analysis system.

The LabSmith Microfluidic Platform

Work in our laboratory has transitioned away from large, precision HPLC pumps and valves to a miniature, highly portable commercial system produced by LabSmith[1]. This system is compatible with 150μm internal diameter inert PEEK[2] tubing (360μm outer diameter), with customised substrate and reactant reservoirs on the millilitre scale, precision micropumps capable of handling microlitre volumes to create flow rates down to $\approx$8 nanolitres per second (500nl/min), and three or four-way switching valves with internal PEEK surfaces. All of these components are fully modular and exchangeable with a common locking ferrule fitting for creating watertight microfluidic connections, and a screw-fit breadboard system for holding the various other components in place.

LabSmith also allows the designer a great deal of control over the function of the system through a fully-fledged scripting language named uProcess™. This provides control over valve switching, as well as pump flow rates and volumes in forward and reverse directions, allowing for complex chains of reactions to be designed, including automated serial

[1]LabSmith, Inc., Livermore, California, USA
[2]Poly Ether Ether Ketone

dilutions to recalibrate sensors over time. Figure 10.6 below demonstrates the system in such a configuration, allowing pumps 1 & 2 to create a dilution according to flow rate, prior to dosing with enzyme from pump 3 with subsequent mixing and detection.

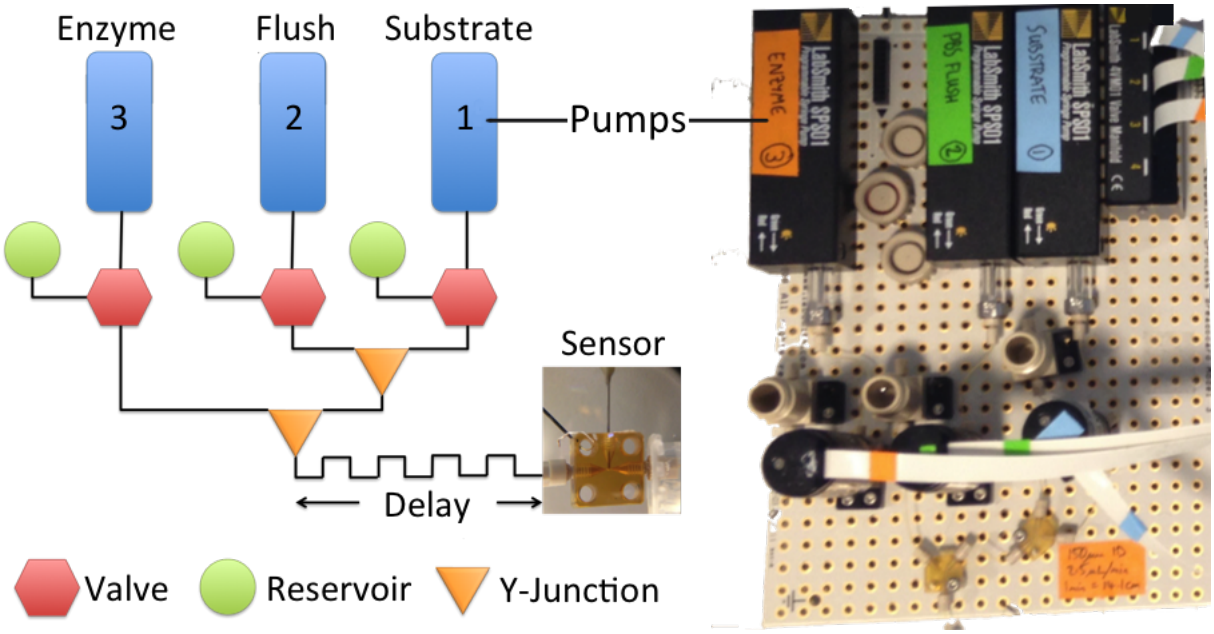


Figure 10.6: Left: Schematic of my LabSmith system configured for serial dilutions **Right:** Photograph of the assembled system

Not only was I able to use existing scripts from lab colleagues, I was also able to create fully customised scripts to match the physical layout and requirements of my particular microfluidic systems. One improvement I brought to the scripts in general use in the laboratory was to de-tension the return spring of the pumps after refilling, which prevented flow spikes on recommencement of forward flow. A second improvement prevented pressure drops in the system during refilling by adjusting the pump timings and by returning the pumps to a neutral setting prior to cessation of flow.

Amperometric Sensors

Amperometry is the technique of measuring the number of electrons consumed or produced by a redox reaction at a certain electrical potential, such as that invented by Leyland Clark (1918 – 2005) in the 1950s for measuring the partial pressure of oxygen in solution (page 273) at a potential of -0.6V to -0.7V vs. Ag|AgCl.

An amperometric sensor comprises three elements – (i) a working electrode to carry out the redox reaction with the substrate of interest, (ii) an auxilary or counter-electrode to balance the other side of the redox reaction, and (iii) a reference electrode to fix the circuit at a stable point in electrical space.

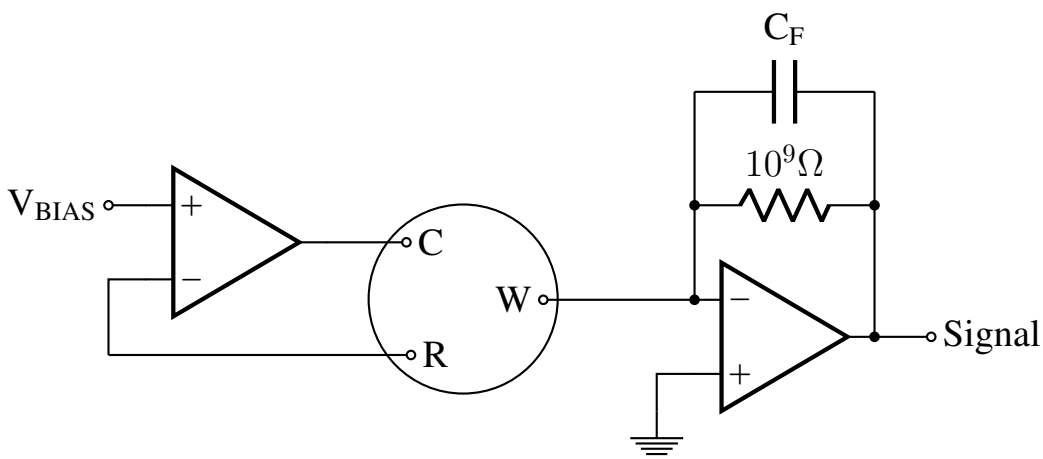


Figure 10.7: A three-electrode sensor circuit comprising a servo amplifier (left) and a transimpedance amplifier with a large gain resistor and stabilising capacitor (right). This shows the reference (R), counter (C) and working (W) electrodes forming the electrochemical cell, and the voltage bias input into the servo amplifier

A potentiostat circuit uses a servo amplifier to automatically adjust the current flow from the counter-electrode to maintain the potential of the working electrode at a fixed point from the reference to control the redox reaction, and is combined with a transimpedance amplifier to measure the current passed by the working electrode as a voltage signal for recording and analysis. This system is shown in Figure 10.7 above.

The transimpedance amplifier must have a suitably high input impedance on the order of $10^{12}\Omega$ to prevent any interference with the redox reaction at the working electrode, and a frequency response to match the expected changes in the system's redox rate with the presentation of new substrate. Similarly, the servo amplifier must have a sufficiently low output impedance and response rate to be able to maintain the stability of potential at the working electrode.

Our laboratory uses a potentiostat designed by a previous PhD researcher, Dr. Chu Wang[167]. This uses the OPA129[1] as the transimpedance amplifier, which has a maximum input bias current of 100 fA, a current noise figure of 0.1 fA/$\sqrt{\text{Hz}}$ and a differential input impedance of $10^{13}\Omega$. In this design, the voltage set point is applied as an inverse voltage to the counter-electrode from the PowerLab data collection system rather than a direct bias at the working electrode, so as to to minimise any possible noise at the inputs of the transimpedance amplifier. The servo part of the circuit uses an OPA140[2] which has a low voltage offset of 120μV, an offset voltage drift of 1μV/°C, a differential input impedance of $10^{13}\Omega$, an output impedance of 16Ω and a gain bandwidth product of 11MHz.

[1]Texas Instruments Inc., Dallas, Texas, USA
[2]Texas Instruments Inc., Dallas, Texas, USA

Needle Microelectrodes

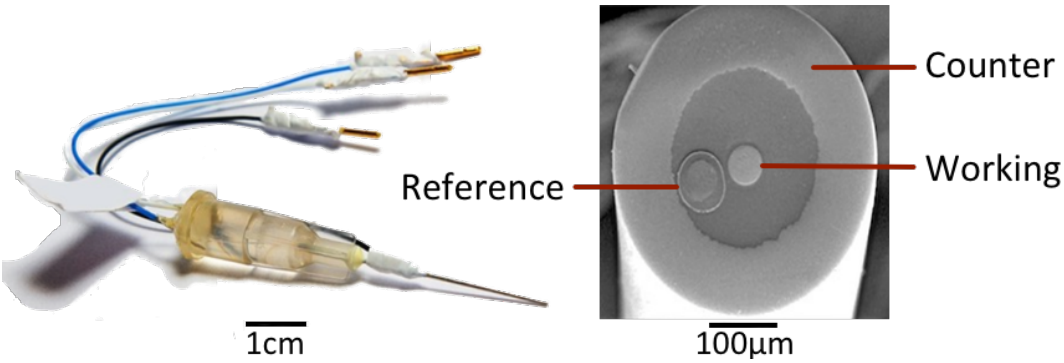


Figure 10.8: Left: A photograph of a needle microelectrode **Right:** A scanning electron micrograph of the tip of a needle microelectrode, demonstrating the working and reference electrodes contained within 27 Gauge stainless steel needle (400 μm outer diameter) which acts as the counter electrode

It was important to be able to develop an amperometric sensor on the same scale as the microfluidic channels within the LabSmith system in order to achieve the goal of detecting the H_2O_2 produced from creatinine. Our laboratory has experience of creating ultra-miniature three-electrode sensors within a 27 Gauge (27G) needle (400 μm in outer diameter)[168] and one such completed 'needle microelectrode' is shown in the photograph and scanning electron micrograph in Figure 10.8 above. I will review the theory and practice behind the design and preparation of our standard 'needle microelectrodes' before introducing my improvements in the following Section.

Whereas much of the research in our laboratory with respect to glucose and lactate sensing employs needle electrodes configured as biosensors, I chose to use the electrode as a plain amperometric sensor, with the enzymes digesting the creatinine in flow rather than at the sensor. This would allow me to perform a greater range of experiments with the more complicated triple-enzyme system required for detecting creatinine, yet leave open the possibility of translating this work into a biosensor at a later stage.

Selecting the Working Electrode Material

Platinum is the material of choice for the working electrode as it is able to oxidise H_2O_2 at low working potentials (+0.7V vs. Ag|AgCl) without becoming oxidised itself, as is the case with gold. The reaction between platinum and H_2O_2 was studied in detail by Simon B. Hall, Emad A. Khudaish and Alan L. Hart in 1997[169], providing valuable insights into the mechanism at the platinum surface which will be briefly explored here. The overall surface reaction is shown equation 10.1.

$$\mathrm{Pt} + \mathrm{H_2O_2} \longrightarrow \mathrm{Pt} + 2\,\mathrm{H^+} + 2\,\mathrm{e^-} + \mathrm{O_2} \tag{10.1}$$

Monitoring the electron transfer from the working electrode via a potentiostat provides a signal that is proportional to the H_2O_2 concentration, and thus the creatinine concentration. However, this reaction appears to deviate from linearity at high concentrations ofH_2O_2 in the fashion of a saturable Michaelis-Menten mechanism. The reaction scheme is currently believed to occur in three steps, the first of which is activation of the platinum surface at a potential of +0.7V vs Ag|AgCl, causing the platinum to form platinum hydroxide ($Pt(OH)_2$) from the oxidation of water, shown in equation 10.2 below.

$$\mathrm{Pt} + 2\,\mathrm{H_2O} \longrightarrow \mathrm{Pt(OH)_2} + 2\,\mathrm{H^+} + 2\,\mathrm{e^-} \tag{10.2}$$

This is the catalytic agent at the surface of the electrode. In the two subsequent steps, platinum hydroxide first adsorbs H_2O_2 and then reduces it, releasing water and oxygen and regenerating platinum metal in the process shown in equation 10.3 below. The electron flow detected by the potentiostat actually corresponds to the electrochemical regeneration of the platinum hydroxide from the platinum metal rather than the decomposition of the H_2O_2, but this point is academic. This decomposition of H_2O_2 is the rate limiting step for the whole process.

$$\mathrm{Pt(OH)_2} + \mathrm{H_2O_2} \rightleftharpoons \mathrm{Pt(OH)_2 \cdot H_2O_2} \longrightarrow \mathrm{Pt} + 2\,\mathrm{H_2O} + \mathrm{O_2} \tag{10.3}$$

The surface availability of catalytic $Pt(OH)_2$ is limiting in this process, and saturation with H_2O_2 will reduce the reaction rate and the sensor's current response. Another feature seen with platinum electrodes is the phenomenon of poisoning, whereby the catalytic sites are permanently deactivated by adsorption of oxygen ($Pt(OH)_2 \cdot O_2$), or protonation of the adsorbed H_2O_2 complex ($Pt(OH)_2 \cdot H_2O_2 \cdot H^+$). Neither of these compounds are able to be regenerated back into platinum metal to continue the reaction process. The relative proportions of these poisoning species changes with the concentration of H_2O_2 in solution, which provides one explanation for the gradual loss of sensitivity over time, and the need to periodically re-condition the electrode. Figure 10.9 shows the evolution of these species as a function of bulk H_2O_2 concentration.

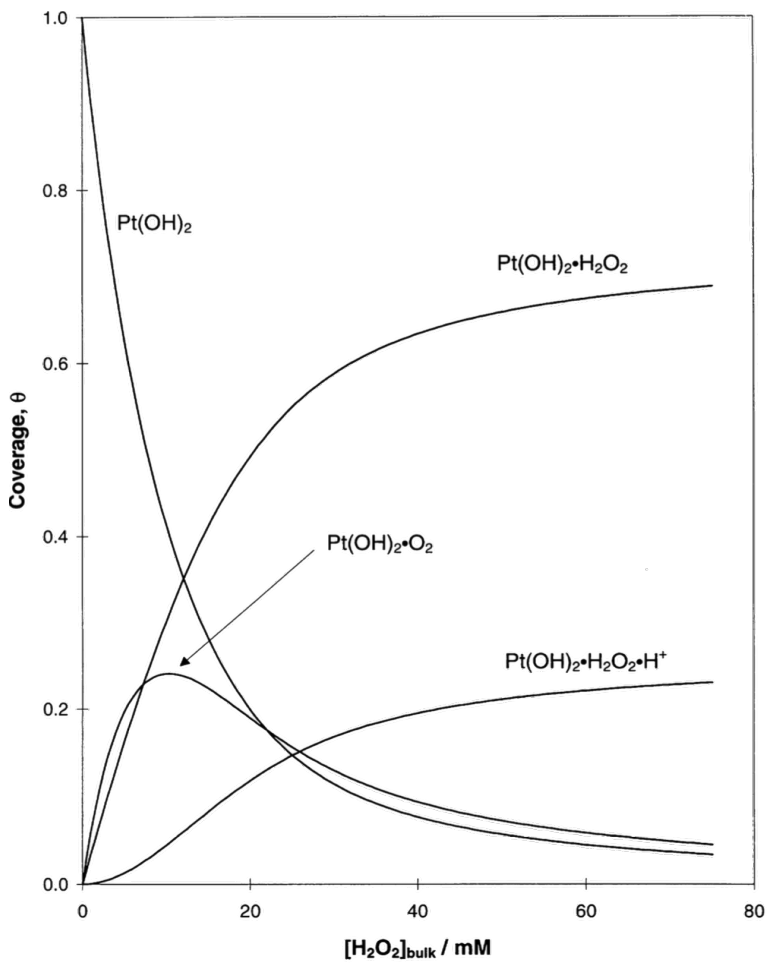


Figure 10.9: The relative surface coverage by different platinum species during hydrogen peroxide oxidation on a platinum surface. $Pt(OH)_2 \cdot O_2$ and $Pt(OH)_2 \cdot H_2O_2 \cdot H^+$ inhibit the detection of H_2O_2. Figure from [169]

Predicting the Response of the Working Electrode

Detection at the small scales of the LabSmith system means using working electrode surfaces less than the internal diameter of the tubing, on the order of tens of microns. Such microelectrodes exhibit different behaviours than larger scale electrodes. In particular, the diffusion field surrounding the working surface of a polished disc microelectrode extends out beyond the diameter of the electrode itself to form a hemispherical surface receiving incoming substrate flux by mass transport. This rate of flux is higher than that of a macroelectrode, allowing the system to achieve large steady-state currents (in the nanoampere range) despite their small size.

The relevant equations governing this behaviour derive from the solution to the convective diffusion equation. These are given below in 10.4 and 10.5 for two different types of microelectrode.

For a flat inlaid disc electrode lying flush with the surface:

$$I = 4nFDrC \tag{10.4}$$

For a hemispherical disc electrode, proud of the carrier surface:

$$I = 2\pi nFDrC \tag{10.5}$$

Where F is the Faraday constant (96,485 Coulombs/mol), D is the diffusion coefficient, r is the radius of the electrode, C is the concentration of the substrate in question, and n is the number of electrons transferred.

Using published diffusion coefficients for ferrocene monocarboxylic acid[170][1] and the two-electron process of H_2O_2[2][171], we can calculate the theoretical maximum performance of our microelectrodes according to the sizes of platinum wire available from commercial suppliers to provide a point of comparison against real-world performance. Ferrocene monocarboxylic acid is a fully-reversible redox couple used to characterise the quality of our microelectrode surfaces through cyclic voltammetry (discussed further on page 341). The results are presented in Table 10.1 below.

Pt Wire Diameter (μm)	1.5mM Ferrocene		1mM Hydrogen Peroxide	
	Flat Disc (nA)	Hemisphere (nA)	Flat Disc (nA)	Hemisphere (nA)
25μm	4.2	6.5	12.5	9.9
50μm	8.3	13.0	25.1	19.7
125μm	20.7	32.5	62.7	49.2

Table 10.1: Calculated maximum currents that can be obtained from platinum microelectrodes by size and redox substrate

Selecting the Reference Electrode

Reference electrodes used in electrochemistry should have a fixed and stable potential in the solution of interest. Many experimental designs have moved away from the classical saturated Calomel (Hg_2Cl_2) electrode to the silver-silver chloride electrode due to concerns over the toxicity of mercury salts. Both reactions take place in saturated KCl solution. The overall redox reactions of each are shown in equations 10.6 and 10.7 below.

[1] 5.73×10^{-10}m^2/sec for ferrocene monocarboxylic acid
[2] 1.3×10^{-9}m^2/sec for H_2O_2

$$\mathrm{Hg_2}^{2+} + 2\,\mathrm{e}^- \rightleftharpoons 2\,\mathrm{Hg}^+ + 2\,\mathrm{Cl}^- \tag{10.6}$$

$$\mathrm{AgCl} + \mathrm{e}^- \rightleftharpoons \mathrm{Ag} + \mathrm{Cl}^- \tag{10.7}$$

The particular electrical potentials (E^0) of these systems versus the 'zero' of the standard hydrogen electrode do not matter, so long as only one stable reference point is chosen for the system. Reference electrodes should ideally have very low impedances at their junction with the sample liquid, so that they are able to pass electrons in order to maintain the charge density in their local vicinity. This prevents polarisation and ensures a stable potential. The Ag|AgCl system comes close to this behaviour. Furthermore, limited polarisation reduces any motion artefact caused by the movement of charge carriers around the electrode relative to the solution[172].

However, Ag|AgCl references are not purely resistive and they do exhibit frequency dependent impedance which rolls off at frequencies above 100Hz[172]. Careful design of the referencing portion of a potentiostat circuit is therefore essential to prevent any high frequency noise from reaching the reference electrode and reducing the overall stability and performance of the circuit.

Standard techniques for producing Ag|AgCl reference electrodes include (i) electrochemical methods such as oxidation of Ag metal in the presence of excess chloride ions, for example hypochlorite bleaches, peroxides or dichromate ions, (ii) creating a sintered plug of finely-divided Ag|AgCl around a silver wire, or (iii) electroplating bare silver wire in concentrated hydrochloric acid. The standard technique in our laboratory involves the electrochemical oxidation of the surface of a 50μm diameter silver wire in the presence of excess Cl^-.

Standard Manufacturing Methodology

Our standard method for manufacturing needle microelectrodes derives from work by the O'Hare group[168], and uses PTFE-insulated platinum/iridium wire in a 90%/10% ratio[1] or an Isonel®-insulated[2] pure platinum wire[3] to create the 50μm diameter working electrode. This is paired with a 50μm diameter PTFE-insulated silver wire and both of these are fed back up a 27G needle. The insulation is removed by briefly applying a lighter flame, before soldering the bared ends of the wires to larger wires for interfacing with the

[1] Advent Research Materials Ltd., Eynsham, Witney, UK
[2] Isonel® is a polyester enamel
[3] A-M Systems LLC, Sequim, Washington, USA

potentiostat. The needle is then back-filled with clear epoxy resin[1].

After curing of the internal epoxy, an electrical connection is made with the barrel of the needle using silver-impregnated epoxy to create the counter-electrode. Following this, the end of the needle is clipped and then abraded to flatness with a series of sandpapers of increasing fineness, beginning with 1200 grit, 2500 grit and finally 4000 grit, before moving on to more careful abrasion with a series of alumina slurries with particle sizes of 1μm, 300nm and 50nm on a fine polishing pad. This is perhaps the most critical stage in the process, and one which I carried out with repeated examination of the polished tip beneath a high-powered binocular microscope to check for surface abrasions or roughness and to achieve a mirror-shine. The precise technique of polishing varies between individuals, but previous work has found that moving in a figure-of-8 pattern with 90° turns between sets provides a good quality finish. I discovered that ultrasonically cleansing the tip between polishes to remove the rapidly-drying slurry generally reduced the appearance of surface scratches and the need to re-polish.

Following this stage, the silver electrode is chloridised to Ag|AgCl using 'Chloridising Solution'[2], a commercially sourced strongly oxidising dichromate solution containing an excess of Cl^- ions. The surface is then rinsed with 1.5M HCl and dH_2O prior to initial characterisation with cyclic voltammetry.

Electrode Characterisation with Cyclic Voltammetry

Cyclic voltammetry (CV) measures the current evolved by an electrode in response to a change in the bias potential over time. There are a multitude of analytical techniques based on various forms of cyclic voltammetry, but we use a voltage sweep from -100mV to +500mV and back at a scan rate of 10mV/s in a 1.5mM solution of ferrocene monocarboxylic acid (ferrocene)[3]. This sweep oxidises ferrocene at the electrode surface in the forward direction, and subsequently reduces it in the reverse direction. Figure 10.10 below demonstrates both the scanning profile and the resultant CV from a well-polished 50μm platinum needle microelectrode, with a peak of 8.6nA vs. the theoretical 8.3nA for a perfectly flat disc.

[1]Robnor Resins Ltd., Swindon, UK. Numbers CY1301 and HY1300
[2]Bioanalytical Systems, Inc., West Lafatette, Indiana
[3]Our stock solution contains 1.5mM ferrocene monocarboxylic acid, with 100mM sodium citrate, 150mM sodium chloride, and 1mM EDTA, in dH_2O

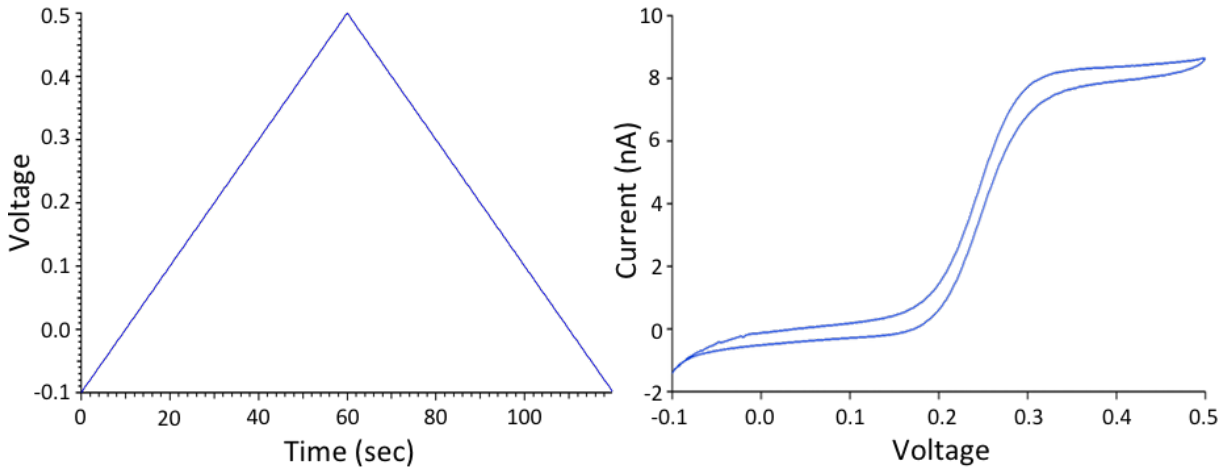


Figure 10.10: Left: Bias potential applied to the cathode **Right:** Cyclic voltammogram of a well-polished 50μm platinum microelectrode, demonstrating a peak current of around 8.6nA

This simple process allows us to detect inadequate polishing by comparing against the expected peak results of Table 10.1 on page 339, as well as examining the angle of the slope, the stability of the oxidation plateau and the separation of the oxidation and reduction curves.

With a well-fabricated electrode, the forward and reverse reactions should be identical, with the slope of the CV (the separation of the oxidation and reduction peaks) determined by Nernstian characteristics $\left(\frac{RT}{F}\right)$ for the single-electron transfer reaction of ferrocene.

As discussed on page 338, the high rate of substrate flux to the electrode surface should match the rate of electron transfer due to the hemispherical diffusion field of the microelectrodes used in our lab, leading to 'steady-state' sigmoidal CVs with plateau regions as the electrode becomes limited only by mass transport and no longer by the potential energy required for electron transfer.

Cyclic voltammetry also allows the detection of any microscopic surface contamination with silver from the neighbouring reference electrode (appearing as peaks around $\pm$100mV), or inadequate HCl washing to regenerate the stainless steel counter-electrode after chloridisation.

Surface Protection with Electropolymerised *m*-Phenylenediamine (*m*PD)

The final step when preparing our needle microelectrodes is to protect the working electrode from contamination and to only allow molecules on the scale of H_2O_2 to reach the surface. This technique has evolved from multiple reports of polymer films used to entrap enzymes by the electrode surface to form biosensors that exist in the literature, including films of

nafion[173], polypyrrole[174], and polyphenol[175].

The most stable and uniform of these are formed by in-situ electropolymerisation. In this way the precise site, rate and thickness of the final film can be controlled. We have found that polymerising *meta*-phenylenediamine (*m*PD)[176] produces reproducible thin films that are closely adherent to the surface of the working electrode and sufficiently dense as to prevent larger interfering redox species from reaching the electrode surface, such as ferrocene or those commonly found in biological systems (ascorbate, urate or paracetamol[1]) whilst still permitting H_2O_2 at a rate sufficient to give good response times (<1 sec).

The method is straightforward. The needle microelectrode is suspended within a 100mM solution of *m*PD in 10mM phosphate buffered saline at pH 7.4, and a voltage of +0.7V (vs. Ag|AgCl) is applied to the working electrode for 20 minutes until the current diminishes to an asymptotically low level. The electrode is then held at 0V for a further 2–5 minutes before being allowed to air dry, followed by rinsing in dH_2O. The quality of the *m*PD layer is then checked with cyclic voltammetry in ferrocene, as described on page 341, wherein a good result is considered to have reduced the magnitude of the signal peak by 95%, with equal oxidation and reduction profiles and no evidence of silver contamination.

Summary

Electrodes fashioned, conditioned, and tested in this way allow my colleagues to detect glucose and lactate levels within injured human brains in real time, using biosensors formed from single stage oxidoreductase enzymes.

I had to refine the system further to enable the detection of serum creatinine at levels of $^1/_5{}^{th}$ to $^1/_{10}{}^{th}$ those of glucose and lactate in addition to a more complex 3-enzyme digestion process. The details of my refinements are covered in the next Chapter.

[1] *N*-(4-hydroxyphenyl)acetamide

Chapter 11

Optimising the Real-Time Creatinine Sensor

This Chapter covers the development and refinement of the sensor system and electronics for amperometric sensing, as well as the development of the microfluidic three-stage enzymatic digestion of creatinine to enable real-time monitoring of creatinine levels.

The Chapter also presents the results of testing the final system in a real-world environment before discussing the potential future direction of research in this area.

11.1 Customising the Detection System

After examining and initial testing with the micro-amperometric setup in our lab, I believed that I could make improvements to the overall signal-to-noise ratio of the system by redesigning the potentiostat, the microelectrode, and the means of placing the electrode into the microfluidic flow stream.

Redesigning the Potentiostat

My initial aim when examining the schematic of our standard potentiostat was to condense the design into a small, portable, battery-powered unit to match the scale of the SmartPipe and the overall concept of the isolated perfusion system. In so doing, I was able to identify potential areas for improvement in terms of the layout, component selection and reference noise abatement.

Layout

I improved upon the hand-soldered two-layer design of the existing potentiostat by using surface mounting and a four-layer board as discussed in Section 6.4 on page 180 of Chapter 6. This provided a large ground plane beneath the sensitive upper component layer to reduce any stray noise, and prevented any leakage from the power lines into the signal traces. Surface mounting also reduces stray capacitances and resistances from excessive use of solder or flux, which is particularly important with sensitive high gain resistors (100MΩ, 1GΩ, and 10GΩ in this design).

Component Selection

The first point I addressed was to source lower-noise components than the OPA129 used for the transimpedance amplifier, and the OPA140 used for the servo amplifier, exchanging them for the LMP7721[1] and the LTC2054HV[2] zero-drift amplifier respectively. The key characteristics are given in the Tables 11.1 and 11.2 below.

[1]Texas Instruments Inc., Dallas, Texas, USA
[2]Linear Technology Corporation, Milpitas, California, USA

Characteristic	OPA129	LMP7721
Input Impedance	$10^{13}\Omega$	$10^{15}\Omega$
Input Bias Current (Typical ‖ Max)	±30fA ‖ ±250fA	±3fA ‖ ±20fA
Noise at 1kHz	$17\text{nV}/\sqrt{\text{Hz}}$	$6.5\text{nV}/\sqrt{\text{Hz}}$
Drift	10μV/°C	-1.5μV/°C
Gain Bandwidth Product	1MHz	17MHz
Power Supply	±15V	±2.5V

Table 11.1: Comparing the LMP7721 against the OPA129. The noise was compared at 1kHz because this was available on both data sheets

Here, the LMP7721 should have a better frequency response than the existing amplifier, with lower noise and drift.

Characteristic	OPA140	LTC2054HV
Input Offset Voltage (Typical ‖ Max)	30μV ‖ 120μV	±0.5μV ‖ ±3μV
Offset Voltage Drift	1μV/°C	0.02μV/°C
Noise (DC – 10Hz)	250nV$_{\text{PP}}$	1.6μV$_{\text{PP}}$
Gain Bandwidth Product	11MHz	500kHz
Power Supply	±2.25V – ±18V	±5.5V

Table 11.2: Comparing the LTC2054HV against the OPA140

With this selection, the difference was less clear cut in terms of overall noise performance, but the LTC2054HV should lead to a more stable system with its auto-zeroing architecture.

Following this selection, I took design cues from the LMP7721 demonstration board from Texas Instruments[177] to ensure adequate shielding of the input stage of the LMP7721, as well as using a further LTC2054HV to actively drive the guard ring mirrored on the upper and lower surfaces of the board. I also placed ferrite beads at the input and output stages of the LMP7721 and LTC2054HV amplifiers to reduce high frequency noise, particularly when connecting or disconnecting external cables to the small 20AWG[1] female pin connectors. The PCB was gold plated to prevent corrosion of the exposed guard traces.

[1] American Wire Gauge, ≈800μm diameter

Power Supply

I realised that running the system from an on-board DC battery could greatly reduce the possibility of any power supply noise interfering with the signal chain, unlike the existing potentiostat which required external ±15V rails. To create the ±2.5V power rails I selected the TPS62205[1], a small +2.5V switching regulator in a surface-mount SOT23 package. For the -2.5V rail I used an LM2687[2] switched capacitor inverter. Both of these devices were specified for high conversion efficiency and low noise.

Unfortunately I had not yet finalised my designs of the SmartPipe's power supply system or multi-function button at this stage, and I wanted to avoid using a microcontroller in this system to reduce the amount of programming I would need to do. As a result, I created a low-battery indicator from a voltage divider and the LT6703[3], a comparator with an internal 400mV reference.

For the charger I selected the MCP73831[4], a fully-integrated device designed for single-cell LiPo batteries. This was not as sophisticated as the BQ24075 later chosen for the SmartPipe and so required an additional FET-based bypass circuit to allow the system to operate from an external DC supply whilst charging the battery.

This design drew a large amount of power and the circuit was only able to operate from a LiPo cell for a few hours between recharging. Longer experiments required the system to be run from a USB supply.

Potentiostat Improvements

The final improvement was to provide a noise bypass loop around the already very stable LTC2054HV used for the servo amplifier. This consisted of a simple two-stage passive RC filter with corners of 16kHz and 1.6kHz to ensure that any high frequency changes at the reference did not feed through to the amplifier. These could have perhaps been reduced by a decade each to achieve even better noise performance.

I kept the existing technique of applying the bias voltage as the inverse magnitude to the servo amplifier rather than the transimpedance stage, but here I implemented an on-board voltage divider from the DC supply. Whilst less noisy than that of the PowerLab, it did require a continuous current draw whilst the system was operational. I also provided the user with a three-way switch to select between internal, ground, and external references (for cyclic voltammetry) in that order. Figure 11.1 below shows the final board design.

[1] Texas Instruments Inc., Dallas, Texas, USA
[2] Also Texas Instruments
[3] Linear Technology Corporation, Milpitas, California, USA
[4] Microchip Technology, Chandler, Arizona, USA

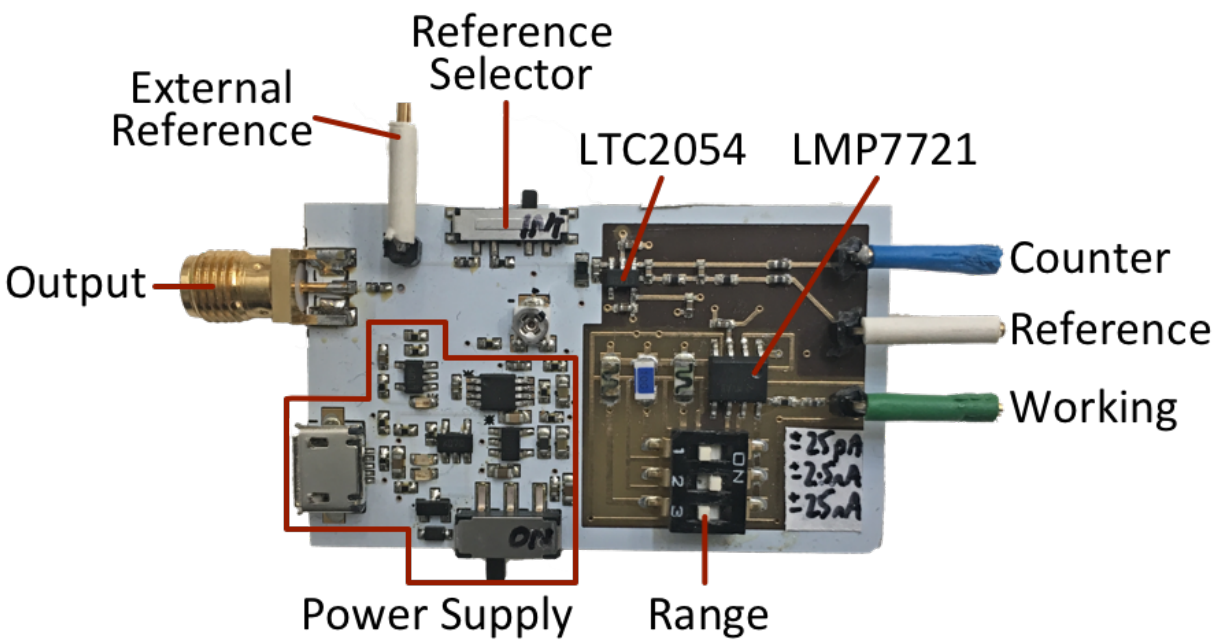


Figure 11.1: Final layout of the PCB from my potentiostat design with the main components labelled. Note the gold plating to prevent corrosion, and the absence of any solder mask or silkscreen from the front-end to minimise stray capacitances

Comparisons with the Existing Design

Figures 11.2 and 11.3 below demonstrate the comparative performance of each transimpedance amplifier stage, using a Keithley 6221 DC Current Source[1] to simulate the current expected from an amperometric sensor. All data was collected using a PowerLab/4SP with a 10Hz low pass filter to reduce the 50Hz noise from the Keithley, and analysed in LabChart. My system was limited to an output range of +2.4V by the +2.5V supply, hence 240pA and 2.4nA.

It is clear that both systems function with high degree of linearity (R^2 coefficients of 0.9999 each against the ideal $x = y$ plot). While Figure 11.3 demonstrates that the performance of my new design is superior to that of the existing system, particularly at low signal levels, both appear to have fixed magnitude gain errors of +3% and -1% respectively when using a 1GΩ transimpedance resistor (1% tolerance).

[1]Keithley Instruments, Inc., Cleveland, Ohio, USA

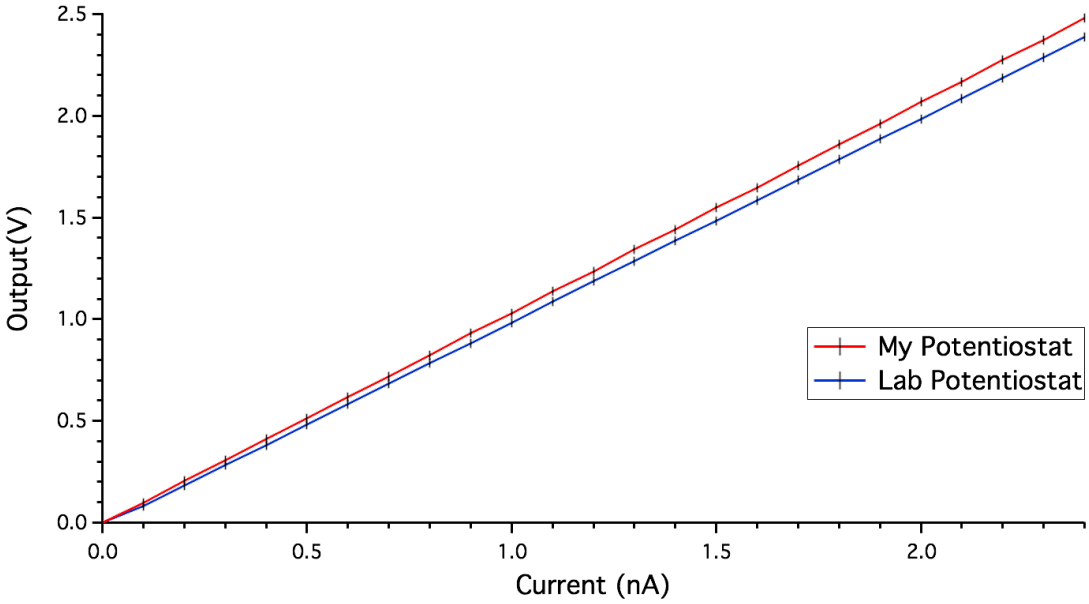


Figure 11.2: Direct comparison between the two systems from 0 – 2.4nA showing high linearity

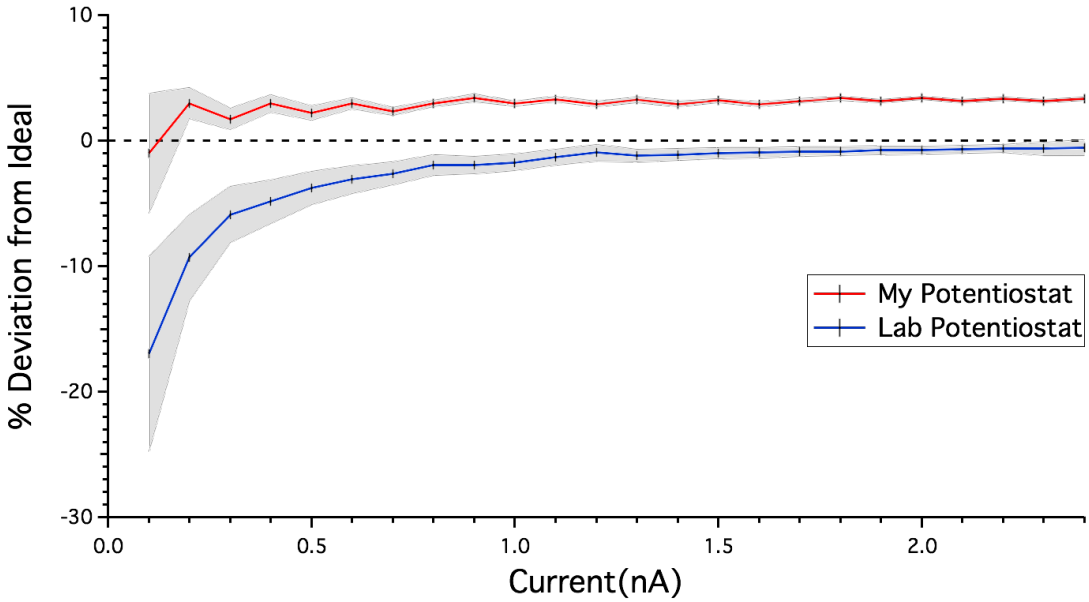


Figure 11.3: Comparing the relative errors between my design and the standard lab system. Shading represents the range of V_{RMS} error. Note the fixed gain error in each system (+3% and -1% respectively), the better performance of my system at low currents, and the appearance of 'steps' arising from the Keithley 6621 DC Current Source

I suspect that this gain error partially results from the 100Ω ballast resistor placed at the input to the LMP7721 creating an additional voltage drop as the current increases, whilst the percentage deviation remains constant. Figure 11.4 below shows the results of testing with currents in the picoampere range, with the lower-tolerance 10GΩ resistor.

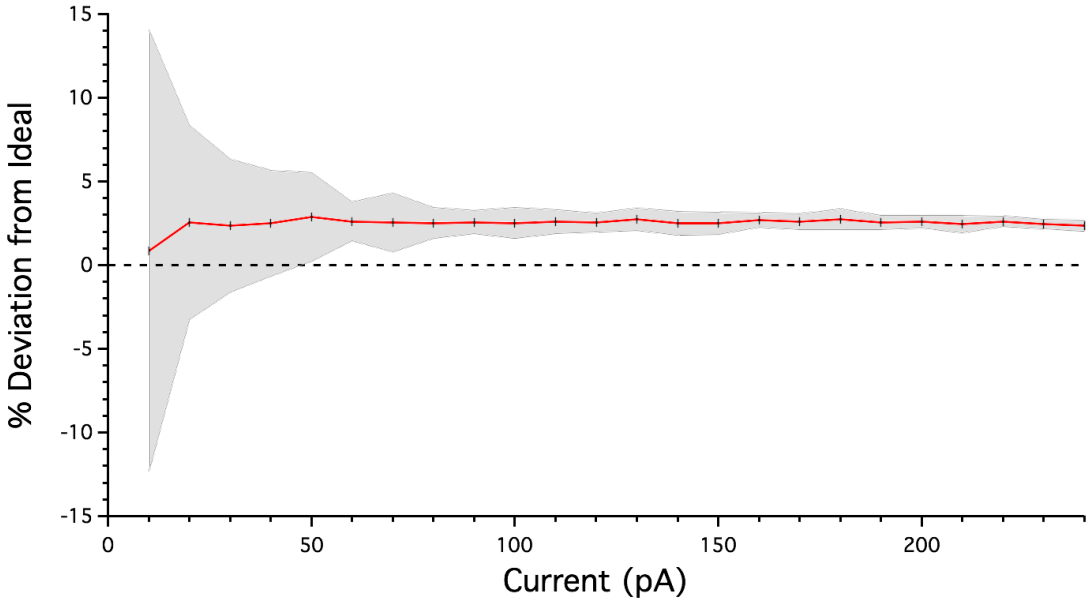


Figure 11.4: Plot of the gain and noise errors (grey shading) between 10 – 240pA

The SNR of my system (Figure 11.5) remains above 30dB down to ≈90pA with the highest gain setting, notwithstanding the inaccuracies in the Keithley current source at low outputs creating the stepped appearance from 50pA – 240pA. My design also had approximately twice the SNR of the original from 100pA – 2.4nA with the $10^9\Omega$ transimpedance gain resistor.

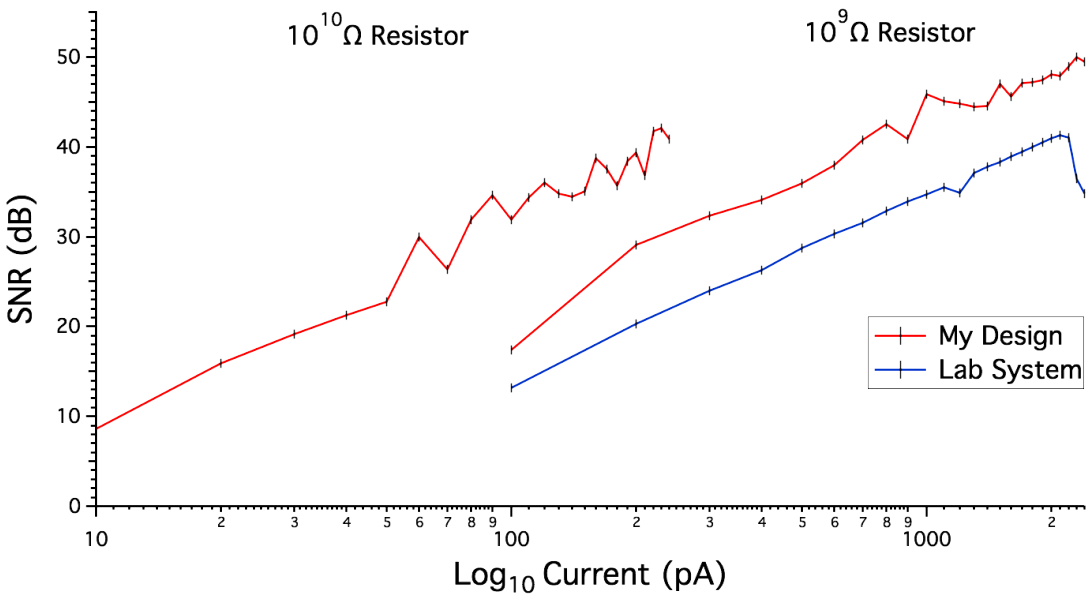


Figure 11.5: The relative SNR performance of each system. The standard potentiostat is in blue, with my design in red. Note the better performance from the higher value TIA resistor down to ≈90pA

Because good SNR performance was more important to the overall detection capability

than gain precision, I felt satisfied that this design had achieved its goals, particularly for low current signals, and could be used to develop the rest of the system.

Improving the Needle Electrode

According to the theory underlying microelectrode behaviour, the noise at the sensor is largely an effect of its impedance[178] which combines resistive and capacitative elements. Microelectrodes minimise their capacitance by generating only small electrical double layers in electrolyte solution, as determined by the Debye length ($\approx 1\mu$m in dH_2O vs. 0.55nm for a 300mM electrolyte solution like plasma). Additionally, their resistances are low because the rates of electron transfer are able to reach the rates of flux into the diffusion field owing to the hemispherical nature of the field emanating from a small point. However, by this same principle, an individual microelectrode is only able to generate small signals due to the low rates of substrate mass transport into a small diffusion field.

For this reason multiple groups have explored the possibility of creating arrays of microelectrodes to gain the low noise benefits of individual point electrodes, yet with greatly increased signal from the parallelisation of diffusion fields within a fixed area[178–180]. These arrays can be micromachined or microdeposited, arranged in strips or points in an ordered or disordered fashion. In all cases, so long as the physical separation of the sensitive surface of the microelectrodes within the array exceeds the dimensions of their diffusional fields (advised as $\approx 10\times$ the radius of an individual element), the SNR of the overall array is greatly boosted in proportion to the ratio of the array area vs. the area of an individual electrode. Figure 11.6 below illustrates the concept of diffusion field coalescence.

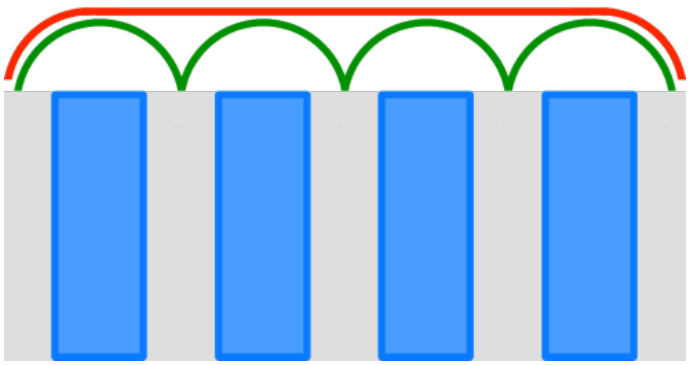


Figure 11.6: The coalesced diffusion fields of a well-proportioned microelectrode array, demonstrating the blue electrodes embedded in a grey matrix, with the individual fields in green and the total effective field in red. Image not to scale

I therefore chose to modify the standard technique for creating needle microelectrodes

to create a random point-array of active working surfaces. This was accomplished by braiding together 8×25μm pure platinum wires insulated by 1μm Isonel® [1] rather than PTFE and feeding this braid into the 27G needle. I had found that PTFE insulation was prone to retract or project from the epoxy encasement, preventing the microelectrodes from functioning normally whereas Isonel® became firmly affixed within the back-filled epoxy and remained flush with the polished surface. After burning away the insulation at the barrel-end of the electrode, the 8 wires were attached to one single electrical contact with either solder or silver epoxy. The rest of the manufacturing process remained unchanged, except that the connecting wires from the counter, reference and working electrodes were twisted to afford a modicum of extra noise resistance in a 'twisted-pair' arrangement so that the potentiostat amplifier could see and reject common mode noise on both the reference and counter electrode connections.

Testing the New Electrode Design

Figure 11.7 shows an image of the 8-point 25μm platinum electrode. Note the random arrangement, which will result in an imperfect, non-uniform coalesced diffusion field.

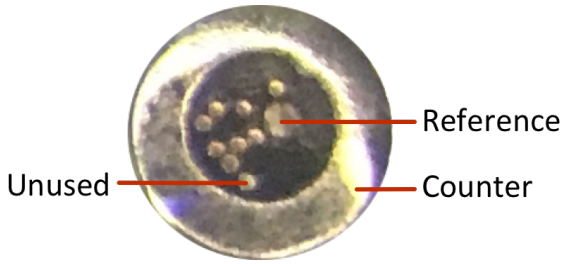


Figure 11.7: Microscope image of the 8×25μm platinum microelectrode array within the 400μm needle microelectrode, containing a ninth, separate electrode for testing

Nevertheless, as shown by Figure 11.8 below, the results were appreciably improved when compared against a standard working electrode tested by cyclic voltammetry in 1.5mM ferrocene solution.

[1] Isonel® is a polyester enamel

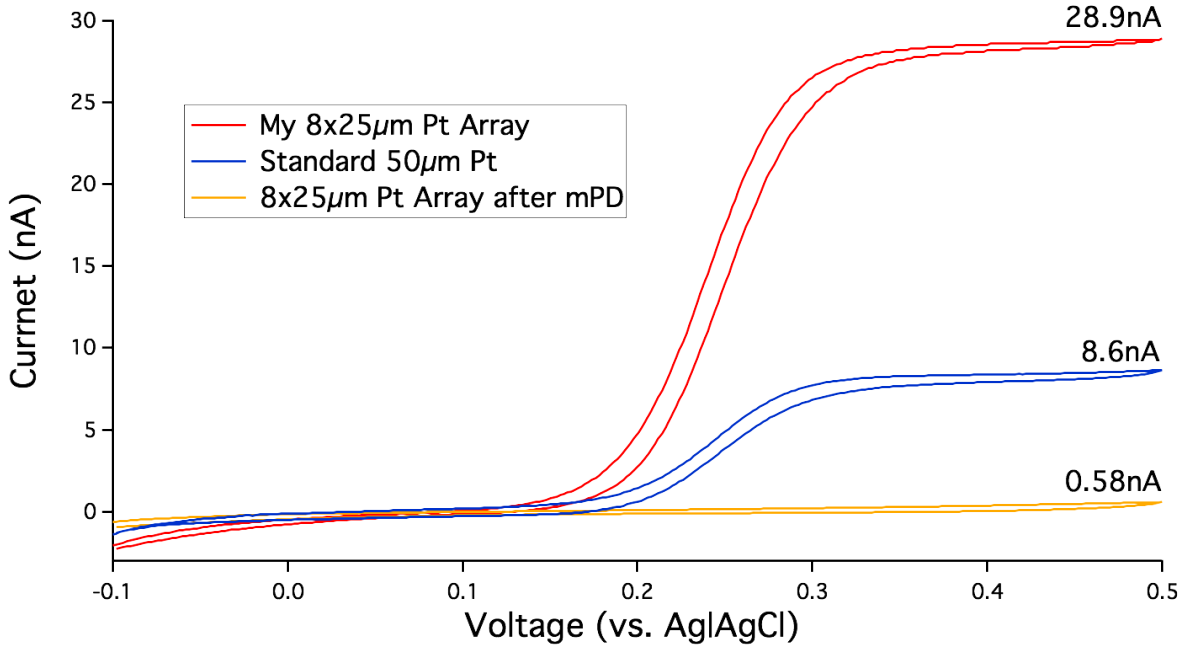


Figure 11.8: Cyclic voltammograms of the 8×25μm microelectrode array vs. the standard single 50μm electrode in our lab standard 1.5mM ferrocene solution, showing the improved signal output of the array. Electropolymerising *m*PD onto the array decreases the current to ≈2% of the original

The array design appears to have a peak current output of 28.9nA, some 3.3 × that of the standard 50μm electrode design, whilst the active surface area is only twice that of the 50μm electrode. Referring back to Equation 10.4 on page 339, the theoretical response for this microelectrode configuration should be 4 times larger. This indicates that despite the electrode proximity apparent in Figure 11.7 (page 352), there is not a great deal of overlap in the diffusion fields of the array and that we are obtaining ≈82.5% of the maximum theoretical signal.

The CV displays a steep slope with low capacitance as shown by the minimal separation between the oxidation and reduction curves. On a practical note, these array electrodes were much easier to polish to a high quality finish than the standard design, requiring only the 1μm alumina slurry following the sandpaper abrasion.

Placing the Microelectrode within the Flow Stream

The existing method of ensuring the working tip of the needle microelectrode resides within the flow stream is to place the needle microelectrodes into a micro-channel embossed into PDMS[1], as shown in Figure 11.9 below. The dialysate enters and exits the chip via push-fit polythene tubing. The problems with this design include leakage of the dialysate from

[1]Polydimethylsiloxane – a transparent, inert silicone elastomer

around the fittings of the inlet and outlet tubing, and the difficulties of introducing the needle electrode just until the tip lies flush with the roof of the analysis chamber to prevent any interference with the flow whilst ensuring good extension of the diffusion layer into the channel.

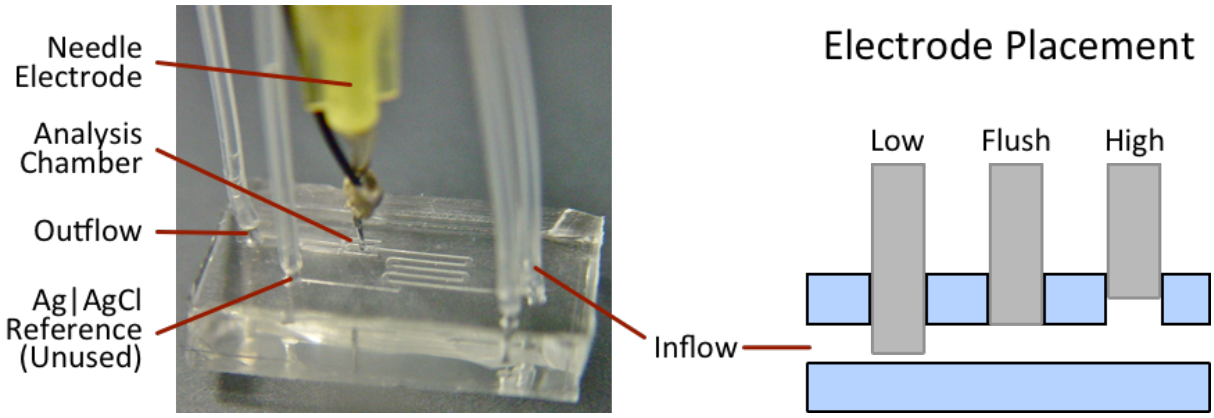


Figure 11.9: Left: A needle microelectrode introduced into the reaction chamber within a custom-designed PDMS chip with a second inlet and external Ag|AgCl reference point designed for droplet microfluidics, but not used in this configuration **Right:** Schematic section through the chip showing the difficulties of introducing the needle electrode into the 50μm high reaction chamber

Instead, I hand-drilled a 400μm channel into a standard LabSmith PEEK T-connector to allow visible placement of the tip within the channel with the aid of a 15$\times$ loupe, whilst retaining the advantage of the modularity and water-tight fittings of the LabSmith system. This is shown in Figure 11.10 below. The tight fit between the outer diameter of the needle and the PEEK drill hole also allowed me to remove and replace the needle multiple times for re-polishing before any leakage appeared.

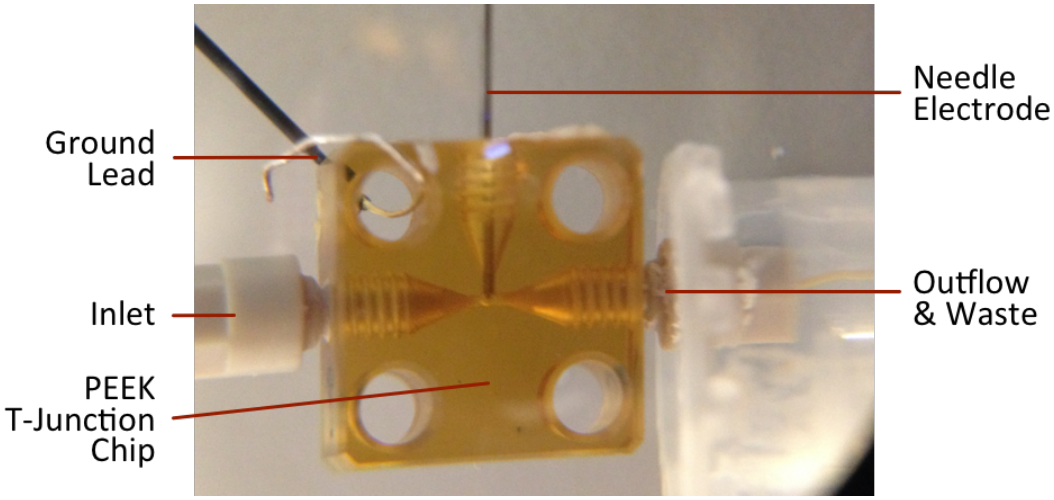


Figure 11.10: Image of the needle electrode placed into a hand-drilled PEEK chip, demonstrating the correct alignment within the flow stream and the modular ferrule fittings of the LabSmith system. Note the small air bubble at the tip of the needle which has entered the system from an imperfect fitting at one of the upstream valves

Summary of Improvements

These redesigns have successfully improved the interface between the electrode and the rest of the LabSmith system, and the SNR of both the potentiostat and electrode array. It should now be possible to push the detection limits of the amperometric system down to those required for quantifying normal serum concentrations of creatinine, which lie between 60μmM – 120μmM. Detection of millimolar urinary creatinine levels could then be achieved by an automated dilution step in the final LabSmith system, using a configuration similar to that shown in Figure 10.6 on page 334 in Chapter 10.

11.2 Optimising the 3-Enzyme System

All experiments used the enzymes creatininase (CNH-311; EC 3.5.2.10; 259U/mg), creatinase (CRH-221; EC 3.5.3.3; 9.18U/mg), and sarcosine oxidase (SAO-351; EC 1.5.3.1; 13.3U/mg), purchased from Sorachim[1] who supply enzymes from Toyobo[2].

This process of refinement took a number of months to complete, exploring the optimal range of enzyme mixtures, buffers and layout of the LabSmith microfluidic system to enable robust detection of creatinine at low concentration.

There were three noticeable trends after reviewing the literature regarding the selection and optimisation of the enzyme reaction. Firstly, the majority of researchers were using

[1]Sorachim SA., Lausanne, Switzerland
[2]Toyobo Co., Ltd., Osaka, Japan

biosensors, with the enzymes embedded in a matrix applied directly to various forms of electrodes. Secondly, there was very little consistency in the specific amounts of enzyme used to create sensors nor the limits of detection derived therefrom. Thirdly, all research on this system over the past 33 years has used phosphate buffered saline (PBS) as the running buffer. Table 11.3 below summarises my literature review with links to the appropriate papers in the bibliography.

Year	Buffer	Conc^n (mM)	pH	Enzymes CA : CI : SO (U/ml)	LOD (μM)	Citation
1983	PBS	50	7.5	6666 : 700 : 253	8.8	[149]
1991	PBS	50	7.5	1.3 : 0.63 : 6.3 †	3	[181]
1991	PBS	50 – 85	7.9 – 8.1	26 : 16 : 9.8		[182]
1992	PBS	100	7.7	300 : 300 : 100	25	[183]
1995	PBS	20	7.5	44 : 35 : 28		[184]
1996	PBS	50	7.5	2200 : 1800 : 1200	0.3	[185]
1996	PBS	50	7.4	2.6 : 3.3 : 3.2 †	30	[186]
1996	PBS	50	7.4	1316 : 200 : 750	10 – 20	[187]
1997	PBS	100	7.0	88 : 70 : 40	1 – 2	[188]
1999	PBS	100	7.5	446 : 656 : 665	200	[189]
2001	PBS	50	7.6	0.3 : 0.3 : 0.3 †	0.8	[190]
2001	PBS	50	7.5	2200 : 1800 : 1200	5.3	[191]
2002	PBS	67 – 335	7.4	4.4 : 0.28 : 0.2 ‡	8	[192]
2003	PBS	100	7.6	§	13.9	[193]
2004	PBS	100	7.0	1185 : 315 : 375	3.2	[194]
2005	PBS	50	7.5	150 : 10 : 18		[195–197]
2006	PBS	100	7.6	§		[198]
2013	PBS	100	7.0	1.29 : 0.26 : 0.42 ‡	12	[199]
2013	PBS	100	7.4	2340 : 1200 : 1600	3	[200]

Table 11.3: Summary of the experimental conditions described in the literature for the three-enzyme amperometric detection of creatinine. CA = creatininase, CI = creatinase, SO = sarcosine oxidase, normalised to U/ml in preparatory solutions, where 1 Unit catalyses the conversion of 1μmol of substrate per minute. † U/cm^2 of electrode ‡ U/electrode § mg of enzyme, unable to perform conversion

Of the papers presented in Table 11.3, only [182] and [183] did not use biosensors, employing instead spectrophotometric and flow-injection-analysis with a sequence of enzyme reaction beds, respectively.

Buffer Selection

PBS is an interesting selection for the running buffer. Whilst the source species of the enzymes in use have changed slightly over the preceding 30 years, creatininases require either Mn^{2+}, Zn^{2+}[201] or Mg^{2+}[149] to function - all divalent cations which are sequestered by PBS. Furthermore, multiple groups have recognised that the optimal pH for the combined enzyme system lies at a more basic pH between 8.0–8.6[149, 181, 183, 186, 192], which happens to coincide with the peak sensitivity of the final enzyme in the chain, sarcosine oxidase.

With respect to the mechanism of H_2O_2 oxidation at the working electrode discussed on page 336 in Chapter 10, Hall, Khudaish and Hart later proposed that the $Pt(OH)_2$ catalytic sites also incorporate phosphate in a pH dependent manner in the form of $H_2PO_4^-$[202]. This appears to promote the decomposition of H_2O_2 and stabilise the surface from one of the poisoning species, $Pt(OH)_2 \cdot O_2$ at phosphate concentrations above 5mM.

They also describe a second catalytic mechanism in the absence of phosphate which has still higher reaction rates, yet requires potentials >458mV vs. Ag|AgCl. This is perhaps overlooked in the literature because of the ubiquity of PBS in these enzymatic studies. This provided reassurance that selecting a different buffer to optimise the reaction rate in flow would not impede the function of the working electrode through the absence of the phosphate mechanism, and that my potentiostat should be able to activate this second mechanism at its bias voltage of 700mV vs. Ag|AgCl.

The final reason for wishing to select a different buffer was the intended use of the system for sampling from either urine or blood. Table 2.1 on page 52 in Chapter 2 shows that urinary pH can be as low as 4.5 (32μmol of H^+) in normal adults. I chose to over-design the system for a pH of 3, to maintain sensitivity in the face of severe ischæmia. The pK_a of phosphate in standard PBS is only 7.2, meaning that a highly concentrated buffer would be required to provide sufficient capacity to neutralise 1mmol of H^+ and maintain the pH of the dialysate within 0.1 unit of pH 8.0. This would require a PBS concentration of 100mM, as demonstrated by using the Henderson-Hasselbalch equation as per 11.1 below, whereas a buffer with a pK_a of 8.0 should only require a concentration of 20mM to resist a pH change of $\pm$0.1 unit.

$$
\begin{aligned}
8.0 &= 7.2 + \log_{10}\left(\frac{\text{Acid}}{\text{Base}}\right) \qquad (11.1)\\
10^{0.8} &= \left(\frac{\text{Acid}}{\text{Base}}\right)\\
\text{Base}(1 + 6.3095) &= 100\text{mM}\\
\text{Base} &= 13.68\text{mM}\\
\text{Acid} &= 86.32\text{mM}
\end{aligned}
$$

Buffering 1mmol of H^+ would change the ratio as follows:

$$\left(\frac{86.32}{13.68}\right) \longrightarrow \left(\frac{85.32}{14.68}\right)$$

Back-calculation with the Henderson-Hasselbalch Equation:

$$
\begin{aligned}
\text{pH} &= 7.2 + \log_{10}\left(\frac{85.32}{14.68}\right)\\
&= 7.964
\end{aligned}
$$

I examined a range of alternate buffers, looking for a suitable buffer with a pK_a of 8.0, low temperature susceptibility, and lack of cation complexation and identified 4-(2-Hydroxyethyl)piperazine-1-propanesulfonic acid (EPPS), an uncommon piperazine-based agent which matched all of these criteria. The better-known piperazine HEPES[1] does complex Zn^{2+}, and the Tris family of buffers (including tricine) are known to complex with Mg^{2+}, Ca^{2+}, Zn^{2+} and occasionally Mn^{2+}[203].

Benchtop tests demonstrated that 50mM of EPPS was able to neutralise a saline solution at pH 3.0 to a final pH of 7.7 when mixed in a 1:4 volumetric ratio with the buffered enzyme solution, versus just pH 7.5 for enzymes in 100mM PBS.

Optimisation Experiments

Previous work in the lab has found that a combination of perfusate flow at 2μl/min and enzyme at 0.5μl/min produce good results. I decided to work backwards from sarcosine oxidase to creatininase, directly testing and optimising each step in turn for the enzyme mixture and pH, prior to performing microdialysis experiments.

Figure 11.11 below shows the results of an initial set of experiments with a single 50μm electrode which I ran prior to creating my 8×25μm electrode, comparing the signal

[1] 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

magnitude of 30U/ml sarcosine oxidase in 10mM PBS versus sarcosine from 25μM to 10mM, confirming my suspicions that basifying the pH to 8.0 would improve the signal. These results are similar for the two-step and three-step mixtures with higher sensitivity at pH 8.0 than 7.5.

These curves were fitted to the underlying data using the Hill Equation functionality within the IGOR Pro[1] graphing software used to produce most of the figures in this section.

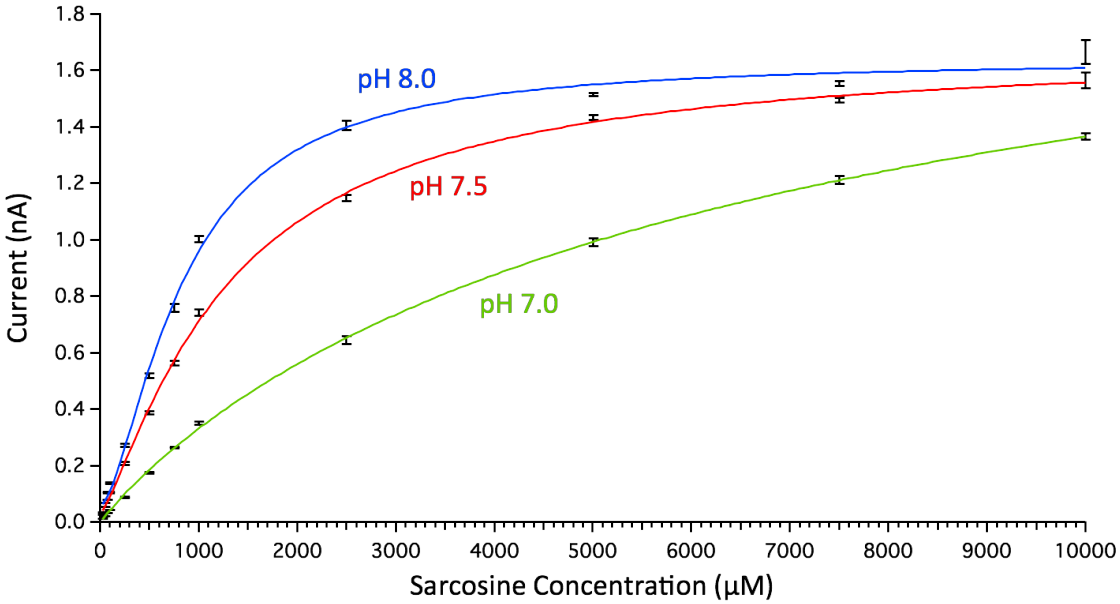


Figure 11.11: Comparing 30U/ml SAO in 10mM PBS from pH 7.0 – 8.0 using a 50μm electrode

A head-to-head comparison of 30U/ml SAO in 100mM PBS at pH 7.5, 50mM EPPS at pH 8.0 and 50mM borate buffer at pH 9.0 provided the results in Figure 11.12 below, using the newer 8$\times$25μm electrode array.

[1]IGOR Pro 6.37, WaveMetrics, Inc., Lake Oswego, Oregon, USA

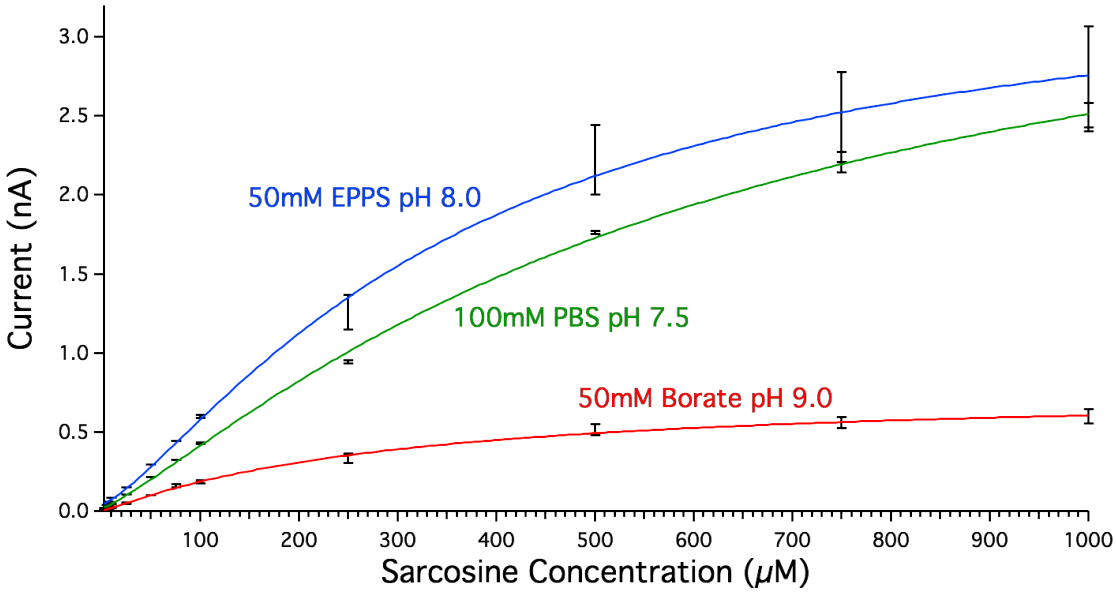


Figure 11.12: Comparing 30U/ml SAO in 100mM PBS at pH 7.5, 50mM EPPS at pH 8.0, and 50mM borate at pH 9.0 with the 8×25μm electrode array, demonstrating a near tripling of the current at 1mM sarcosine in EPPS

The broadening of the standard deviation in the EPPS signal as the concentration progresses was possibly due to a fault with the substrate pump and valve which also appeared in later experiments, leading to their replacement. Figure 11.13 below illustrates the contours of the raw data from which these results were derived. Note the clean and consistent steps obtained during the earlier experiments at lower concentrations, followed by the spike-and-decay pattern and narrow steps for 1000μM sarcosine in 50mM EPPS at pH 8.0.

When the higher concentration experiments were repeated a further 2–3 times in order to derive the mean and standard deviation between runs, further inconsistencies appeared in the signal levels.

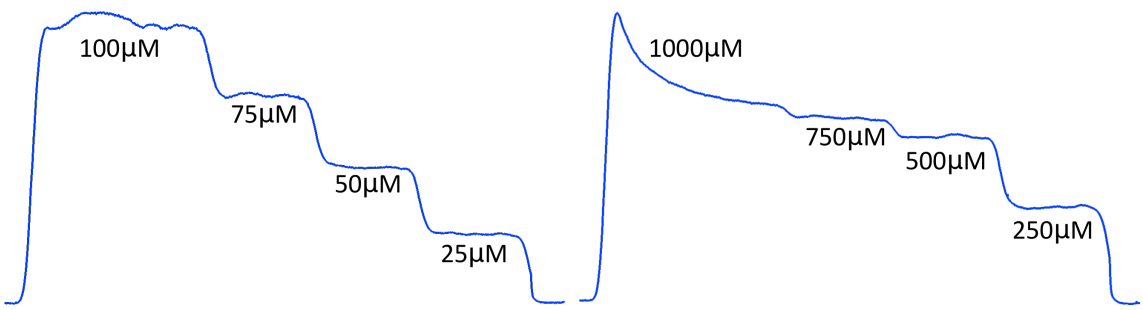


Figure 11.13: Raw signal contours from serial dilution experiments of 30U SAO in 50mM EPPS at pH 8.0 vs. 100μM and 1000μM sarcosine

Figure 11.14 below shows the stepped profiles of these serial dilution experiments with the enzymes in different buffers, demonstrating the clear results obtained in PBS and

EPPS versus those from Tris and borate buffers, confirming their unsuitability for this system. Interestingly, the contour of the Tris buffer system resembles that of the non-functioning EPPS system in Figure 11.13, above. However, unlike the EPPS system, the Tris experiments immediately followed and preceded further experiments in PBS which were of normal contour. This implies that the Tris system is somehow sequestering sarcosine or inhibiting the active sites of the SAO enzyme, leading to the same appearance as a substrate-enzyme delivery mismatch in the higher-concentration EPPS system.

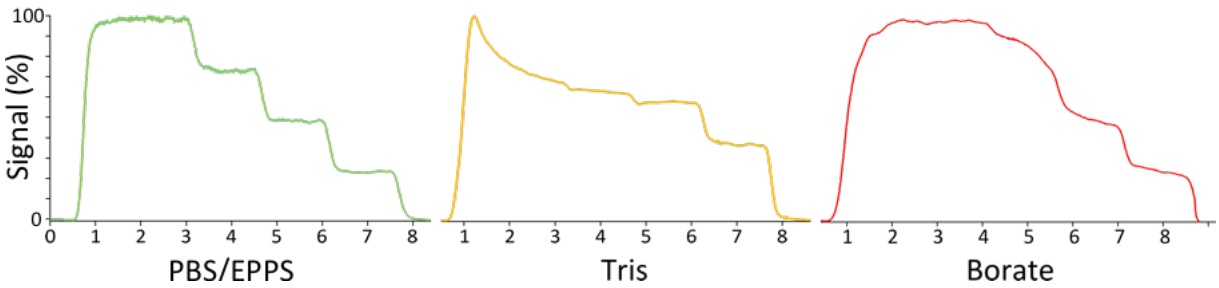


Figure 11.14: Standardised current response profiles obtained from serial dilution experiments of 30U/ml SAO vs. 100uM sarcosine in various buffers, confirming the unsuitability of Tris and borate buffers for this system. Times in minutes

Thereafter followed a series of experiments to examine the time profile of the response curves to various mixtures of enzymes to achieve the maximal response in the shortest time, beyond which minimal improvements could be seen. This would indicate that the enzyme ratios were no longer limiting, merely the amount of enzyme. I decided to limit the total enzyme content of the system (in weight/volume) to that of serum albumin (400mg/ml), but this could be pushed further in later developments. I was mindful of the possibility of encrustation within the microfluidic system, as well as increased viscosity and interference with mixing and substrate diffusion at higher protein concentrations on these scales.

All experiments were carried out with a reservoir of 100 μM substrate in normal saline at pH 3.0 into which the enzyme mixture was added in the intended 1:4 volumetric ratio and then pumped past a sensor at 2.5 μl/minute to reproduce the total flow of the final system. The enzyme mixtures were buffered in 50mM EPPS at pH 7.5, 8.0 and 8.5. The extensive series results will not be reproduced here, except for Figure 11.15 below which was one of the final experiments wherein the SAO and creatinase content had been optimised for 100μm creatine, and this experiment was now attempting to ascertain the optimal amount of creatininase for 100μm creatinine in normal saline at pH 3.0.

Note how increasing the pH from 8.0 to 8.5 was the equivalent of doubling the amount of creatininase content from 300U/ml to 600U/ml (blue vs. red lines), and the increased response with a mixture of 600:300:60 at pH 8.5. The final mixture chosen for the microdialysis experiments was 600:300:60 in 50mM EPPS at pH 8.0, but this experiment

raised the possibility of using an alternative buffering agent with a higher pK_a around 8.5 in future, such as HEPBS (pK_a of 8.3)[203].

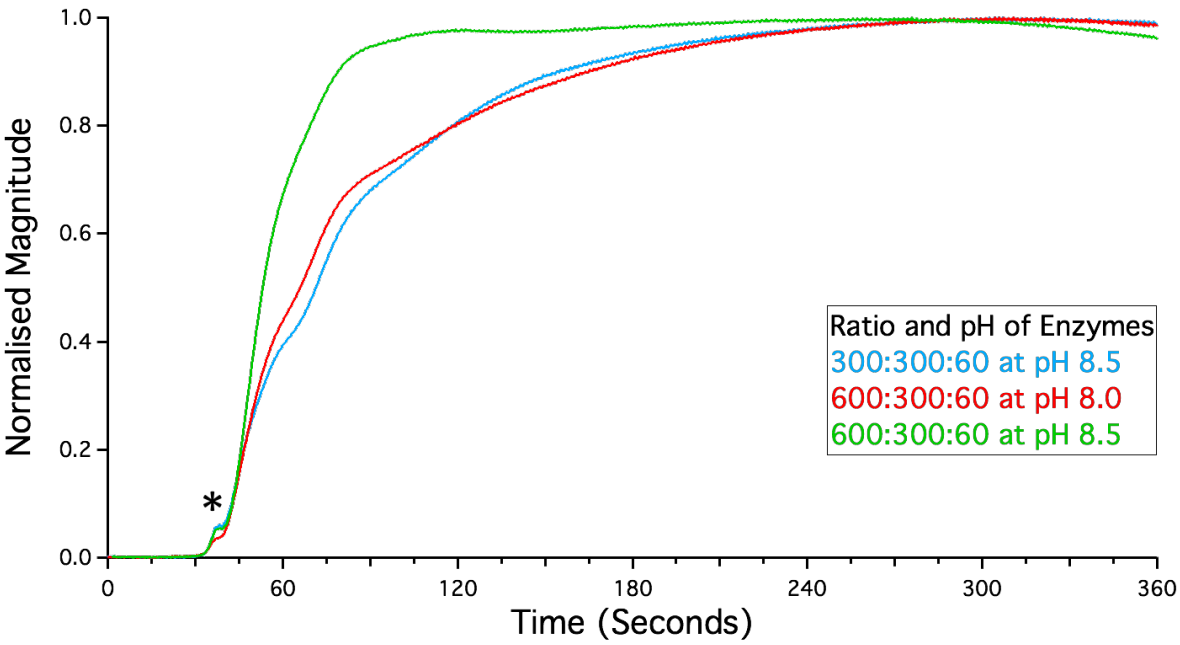


Figure 11.15: One of the final enzyme optimisation experiments demonstrating the normalised time evolution of the signal from the enzymatic digestion of 100μM creatinine. All enzyme amounts in Units/ml. Note the small perturbation ($*$) caused by a leading edge of unbuffered NaCl at pH 3.0

Table 11.4 below presents a collection of T_{90} levels[1] obtained by this experimental method at pH 8.0, demonstrating the evolution of the mixture.

SAO (U)	T_{90} (sec)	CRH : SAO (U)	T_{90} (sec)	CNH : CRH : SAO (U)	T_{90} (sec)
15	138	150 : 60	145	150 : 300 : 60	195
30	73	180 : 60	104	300 : 300 : 60	154
60	28	300 : 60	77	600 : 300 : 60	135

Table 11.4: Results of enzyme optimisation experiments at pH 8.0 in order to achieve minimum T_{90} levels. The reaction time of the final mixture is highlighted in green

From these results I decided to implement a 3 minute delay between the Y-junction feeding the enzyme into the dialysate, and the sensor, to ensure maximum sensitivity by providing adequate mixing and reaction time. This was implemented as a hand-fashioned loop made from the same PEEK tubing used for the rest of the system and fixed in form with epoxy, as shown in Figure 11.16 below.

[1]Time to reach 90% of maximum, measured from the beginning of the upstroke

Figure 11.16: The 42-cm long, 3-minute delay loop to ensure adequate mixing and reaction time between the junction of the dialysate and the enzyme stream, and the sensor

Microdialysis Experiments

With the enzyme quantities and buffer optimised for detecting creatinine at levels of 100μM, I moved to test the system in a simulated final setting with microdialysis. Here, a clinical-grade CMA 70 microdialysis probe [1] designed for deep tissue sampling, with a membrane surface area of 18.8mm^2 and cut-off of 20kDa was suspended in well-stirred T1 solution (an extracellular fluid analog)[2] to which was added aliquots of creatinine in a standard-addition methodology. T1 was also used as the perfusate, delivered at 2μl/min by a Harvard Apparatus PHD 2000 programmable infusion pump[3], with the dialysate returning into the Y-junction of my LabSmith board to mix with the buffered enzyme mixture flowing at 0.5μl/min, followed by the delay loop (Figure 11.16 on page 363) and sensor. From these results it was possible to build a calibration curve for the system, which fit the Hill Equation for enzyme kinetics with a K_m of 2.3mM ($\pm$1.3mM), V_{max} of 2.9mM ($\pm$1.0mM) and rate constant of 0.96μM/sec ($\pm$0.05μM/sec). Interestingly, the system's K_m value encompasses that of sarcosine oxidase (K_m of 2.8mM), but not creatinase (4.5mM) or creatininase (32mM) which could indicate that this is the rate limiting step, perhaps even due to the availability of oxygen in solution ($\approx$250μM).

The same setup was then used for standard addition experiments in well-stirred defibrinated horse blood[4] to prove that it was possible to detect micromolar quantities of creatinine in a biological fluid. The results are given below in Figure 11.17.

[1] M Dialysis AB, Stockholm, Sweden
[2] Our stock solution of contains 2.3mM calcium chloride, 147mM sodium chloride, and 4mM potassium chloride in dH_2O
[3] Harvard Bioscience Inc., Holliston, Massachusetts, USA
[4] TCS Biosciences Ltd., Botolph Claydon, Buckingham, UK

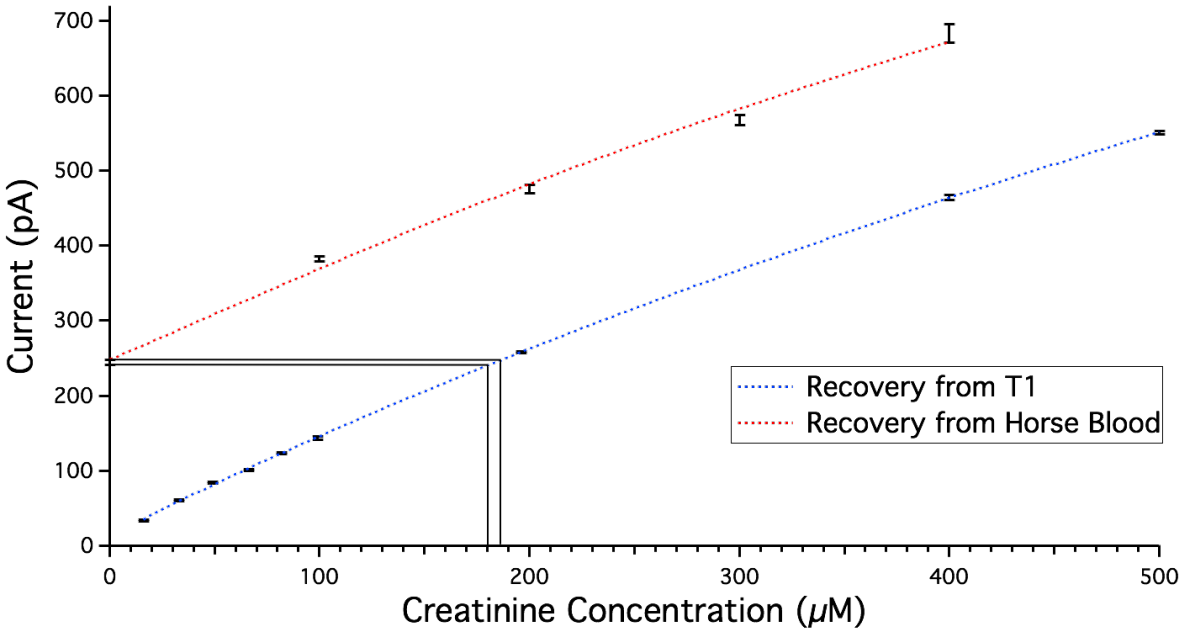


Figure 11.17: Creatinine calibration curve for microdialysis at 2μl/min, obtained by standard addition in well-stirred T1, with a parallel sampling curve from well-stirred defibrinated horse blood. Both curved obtained by auto-fitting to the Hill Equation, with error bars from the perturbations in the continuous raw data

The results obtained in T1 show that this microdialysis setup, the first of its kind, is a sensitive and low-noise method for measuring creatinine, with a limit of detection of 4.3μM and tested upper range of 500μM. The K_m of the curve indicates that this method could be useful up to levels around 2mM after further testing, providing a broad useful working range. Furthermore, the microdialysis sampling methodology only had an estimated recovery of 40%, meaning that improving recovery could push the limit of detection down to $\approx$2μM.

The results in well-stirred horse blood show that the horse had a basal creatinine level of 180μM – 186μM. This is just beyond the upper range of normal for a horse (100μM – 160μM), but we do not know the muscle mass or gender of the horse from which this was obtained, nor their exercise status, whereby levels can rise to $\geq$200μM[204]. It is also possible that the sample was slightly hæmolysed, with erythrocyte creatine feeding into the enzyme cascade (see Table 11.5 on page 366 for a discussion of this point). The broader standard deviations of these results no doubt come from a combination of convection effects and excluded diffusion paths due to red cell mass, altering flux across the dialysis membrane in a chaotic fashion.

Stability Testing

In order to test the long-term stability of this microdialysis system, I suspended the probe in a well-stirred pot of T1 to which was added an amount of creatinine to bring the total concentration to 100μM. The normalised results in Figure 11.18 show that the system remains responsive over a 12 hour period, with the sensitivity falling to a band between 50–60% of the original signal after 9 hours (equivalent to 250pA), but remaining constant from that point onwards. The increase in noise from the $11\frac{1}{2}$ hour mark was the result of colleagues coming in to work in the morning. Spectral analysis showed three major noise peaks – one at 50Hz from the power supply, a second one at 13Hz possibly from the magnetic stirrer, and a much slower 0.2Hz sinusoid superimposed visible over the entire dataset which could reflect convection within the stirred liquid or the screw drive of the Harvard Apparatus PHD 2000 pump.

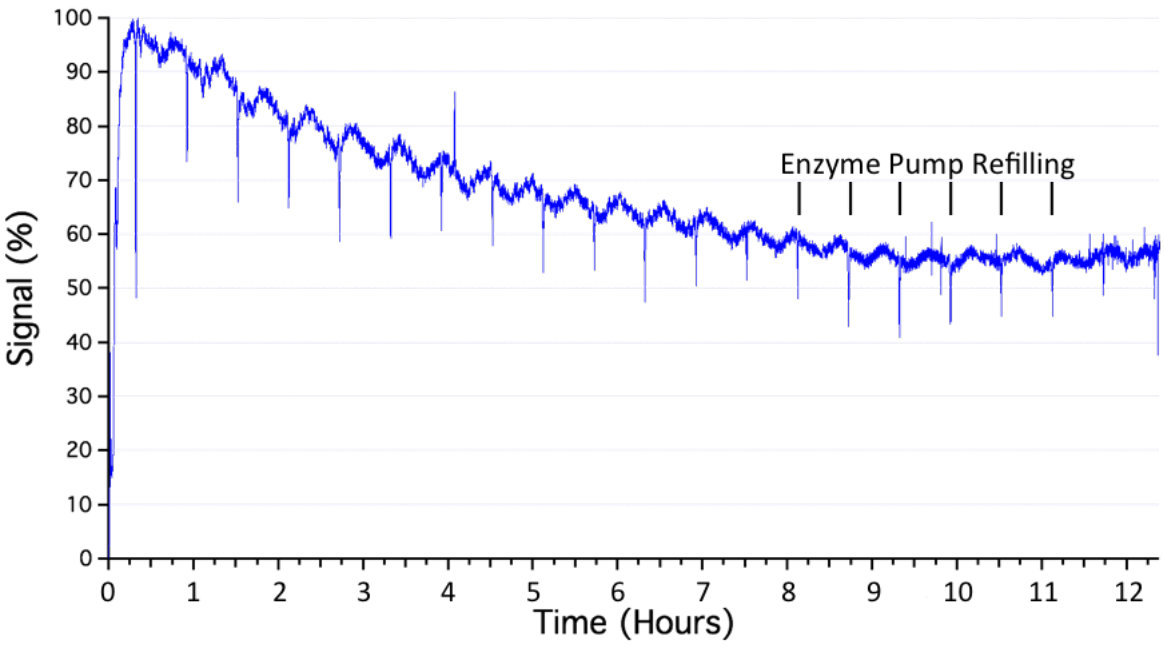


Figure 11.18: Testing for stability of the microdialysis sampling system over a 12 hour period. Note the spikes from the enzyme pump refilling every 40 minutes (20μl at 0.5μl/min). The experiment terminated just beyond the 12 hour period when the enzyme reservoir was exhausted

Interference Testing

At a bias voltage of 700mV versus Ag|AgCl, the working electrode is able to oxidise other chemicals often found in blood such as paracetamol, uric acid, and ascorbate, but these should be prevented from reaching the electrode surface by the polymerised mPD layer. The three-enzyme system will also be able to generate H_2O_2 from sarcosine and creatine.

The levels of these common interferents are presented below in Table 11.5.

Interferent	Concentration	Citation
Ascorbate	164μM †	[186]
	228μM †	[190]
Uric Acid	916μM †	[190]
	42 – 744μM ∗	[205]
Paracetamol	164μM †	[186]
	264μM †	[190]
Sarcosine	0.6 – 2.76μM ∗	[206]
Creatine	25μM ∗	[207]
	38.2 – 68.7μM ∗	[208]

Table 11.5: † Levels used in sensor testing
∗ Documented normal range in serum

I did not test for creatine and sarcosine interference in the final system because the endogenous levels of sarcosine are in the low micromolar range, and those of creatine should only cause problems in the event of extensive hæmolysis, as the majority is intracellular. These two substances could also be accounted for by pre-treatment, background subtraction, or a parallel sampling pathway with a different enzyme mixture.

Figure 11.19 below presents the results of interference testing with the addition of the target substance into a well-stirred container of T1. The ascorbate and paracetamol were added in amounts exceeding those in the literature.

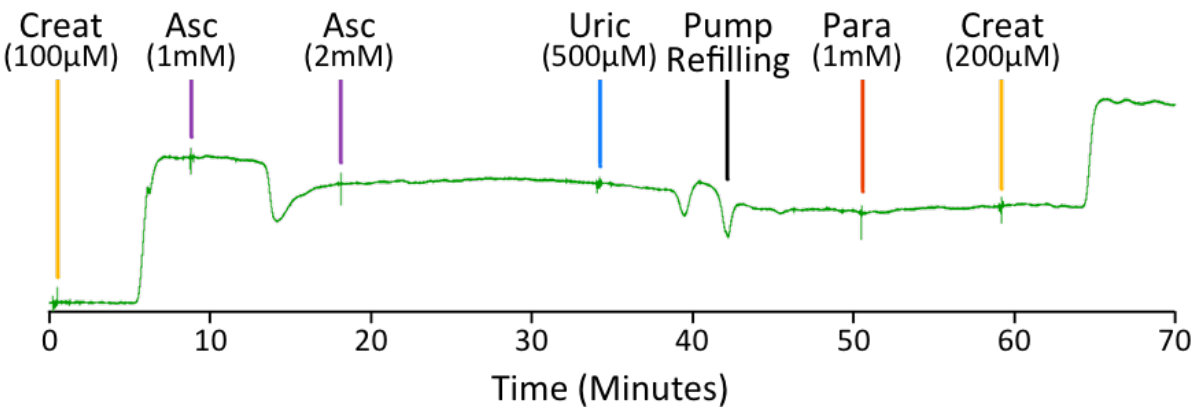


Figure 11.19: Testing interference by ascorbic acid (Asc), uric acid (Uric), and paracetamol (Para). Values below labels are the running total concentration. The concentration of uric acid was estimated from its maximum solubility in water at 20°C

Note the response to the first addition of ascorbic acid but not the second, and a similar response to the addition of uric acid, prior to the pump refilling. These may have been due

to a temporary reduction in recovery as the probe tip came into contact with the inside of the small glass sample pot used for the experiment. There is no apparent interference from the second addition of ascorbate, nor paracetamol, nor any interference with the response to a second aliquot of creatinine to bring the total concentration to 200μM.

It is interesting to note that the molecular cut-off of the polymeric mPD layer must therefore lie somewhere between 34 Da[1] and 151 Da[2].

Despite these good results, I also realised that there could be a different way to discount the effects of any potential interferents in the system.

11.3 Measuring Creatinine Clearance

Problems with measuring absolute magnitudes of responses include the need to continuously account for drift in the sensitivity and offset of the sensor as the working electrode becomes poisoned by H_2O_2, coated with protein, or the reference degrades. There is also a need to account for any potential interferents in the system which may give factitious results, as discussed in the previous section.

I realised that it should be possible to construct a test for the creatinine clearance itself, deriving the renal function as a rate constant rather than measuring absolute concentrations and thereby avoid all potential concerns over interferents and sensor drift, so long as the creatinine remains detectable above background. If we consider the closed loop perfusion system should contain no endogenous creatinine, it should be possible to add a known quantity of creatinine to the circulating volume at regular intervals and monitor for the decay rate as it is filtered into the urine by the working kidney with first-order kinetics. At levels above failure, the clearance should reflect the GFR, as the contribution by active tubular secretion is minimal.

[1]The molecular weight of hydrogen peroxide
[2]The molecular weight of paracetamol – the smallest of the three interferents tested

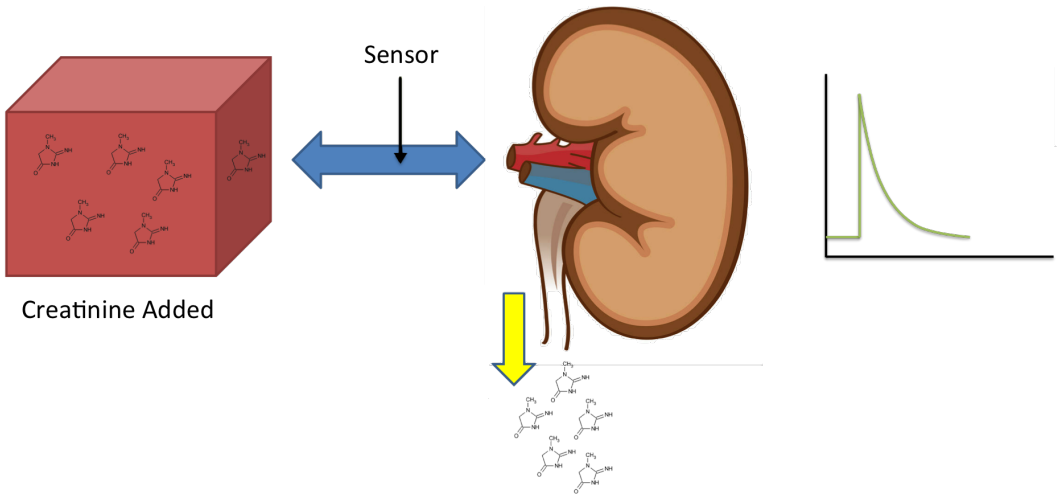


Figure 11.20: Demonstration of the principle of measuring clearance in a closed-loop system by introducing creatinine and monitoring the decay rate as it is filtered by the kidney

I therefore constructed a series of experiments to simulate different creatinine clearance rates for known quantities of creatinine in T1 during continuous microdialysis sampling. For example, a clearance rate of 100ml/min would bring a 1 litre sample circulating at a rate of 1/minute[1] to half of its original concentration in five minutes. This clearance can be simulated by steadily doubling the volume of a 2ml sample containing a known quantity of creatinine over five minutes, or at 400μl/min, using the setup in Figure 11.21 below.

I chose to recreate the clearance rates of kidneys in various states of dysfunction, from CKD1[2] to CKD4, with clearances of 100ml/min, 75ml/min, 50ml/min and 25ml/min respectively. Table 2.2 on page 72 in Chapter 2 first introduced the correspondence between the GFR and the stages of CKD. Note that the signal decay rate during the 12 hour stability test shown in Figure 11.18 on page 365 would be the equivalent to a clearance rate of 2ml/min.

[1]Equivalent to the blood circulation rate of a normal adult human (5 litres of blood at 5l/min)
[2]Stage 1 Chronic Kidney Disease

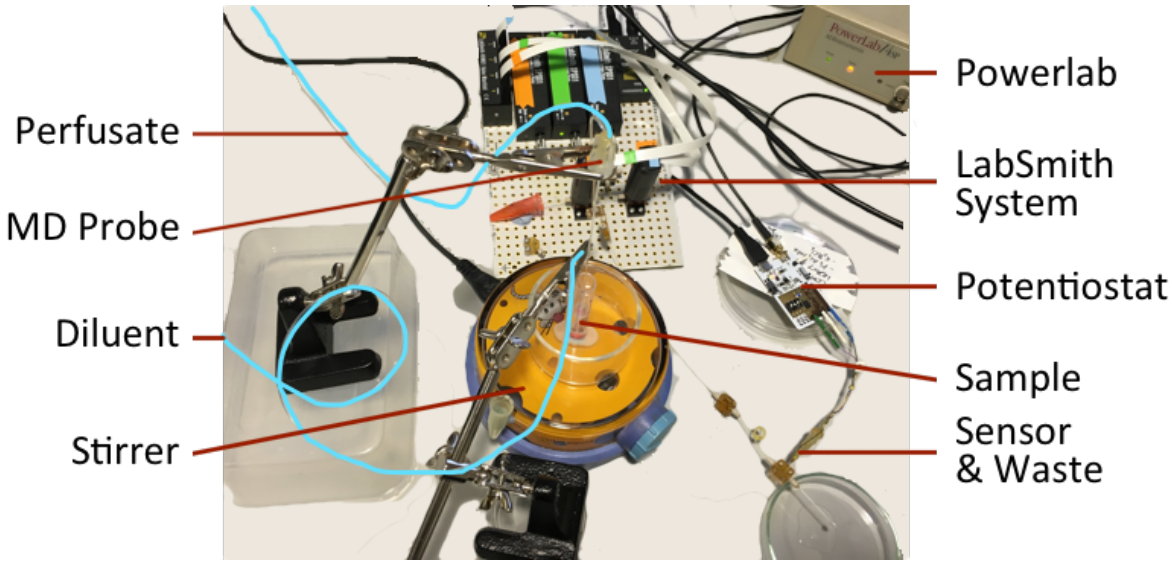


Figure 11.21: Experimental setup to mimic creatinine clearance by sample dilution. The head of the microdialysis (MD) probe is held by the crocodile grip, allowing the probe tip to lie in the centre of the stirred sample below

Figure 11.22 below shows the results of this dilution testing for three different concentrations of creatinine (100μM, 200μM and 300μM) at a simulated clearance rate of 100ml/min.

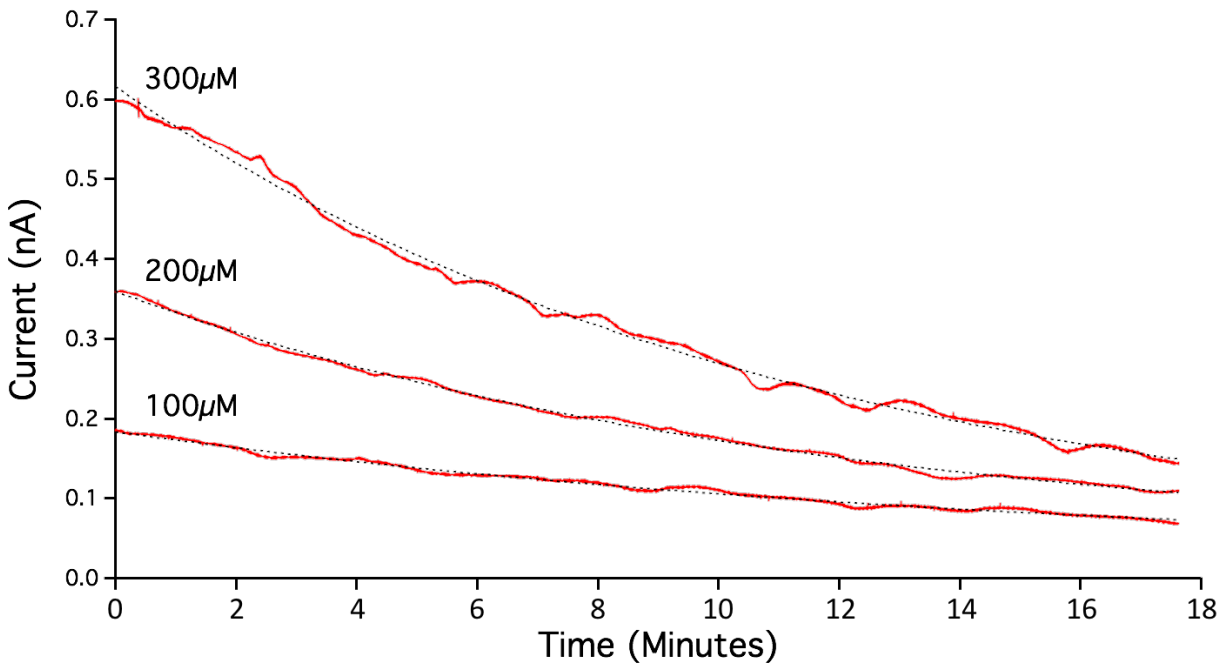


Figure 11.22: Simulating 100ml/min creatinine clearance with different levels of creatinine in solution. The dotted lines show the exponential curves from which the rate constants and half-lives were derived

The results for the 200μM and 300μM experiments were very similar, with time constants

giving half lives of 686.8±1.3 seconds and 679.9±1.6 seconds, respectively. The half life for the 100μM sample was much higher at 895.2±4.0 seconds. It is worth noting that the decay curves are reminiscent of those described by the Albery equation[209], indicating the variability of the supply of substrate to the electrode in the dialysate is perhaps the root cause of these experimental errors, as I did not control for probe placement, stirring rate nor temperature. Improved stirring would be expected to increase recovery in this setup, as would an apparatus to ensure the probe tip did not come into contact with the vessel wall.

Any sensor offset was subtracted from the original data prior to deriving these curves, and therefore any apparent offset that remains is a result of not allowing the experiments to run to completion, wherein each curve would asymptotically approach a current of 0nA.

Figure 11.23 shows the follow-up experiment with 200μM standards with dilution rates selected to simulate different levels of CKD. Each trace has been normalised and time-aligned to begin at 100%, and a log-linear scale has been used to emphasise the different decay rates observed and simplify calculations of the relative half lives.

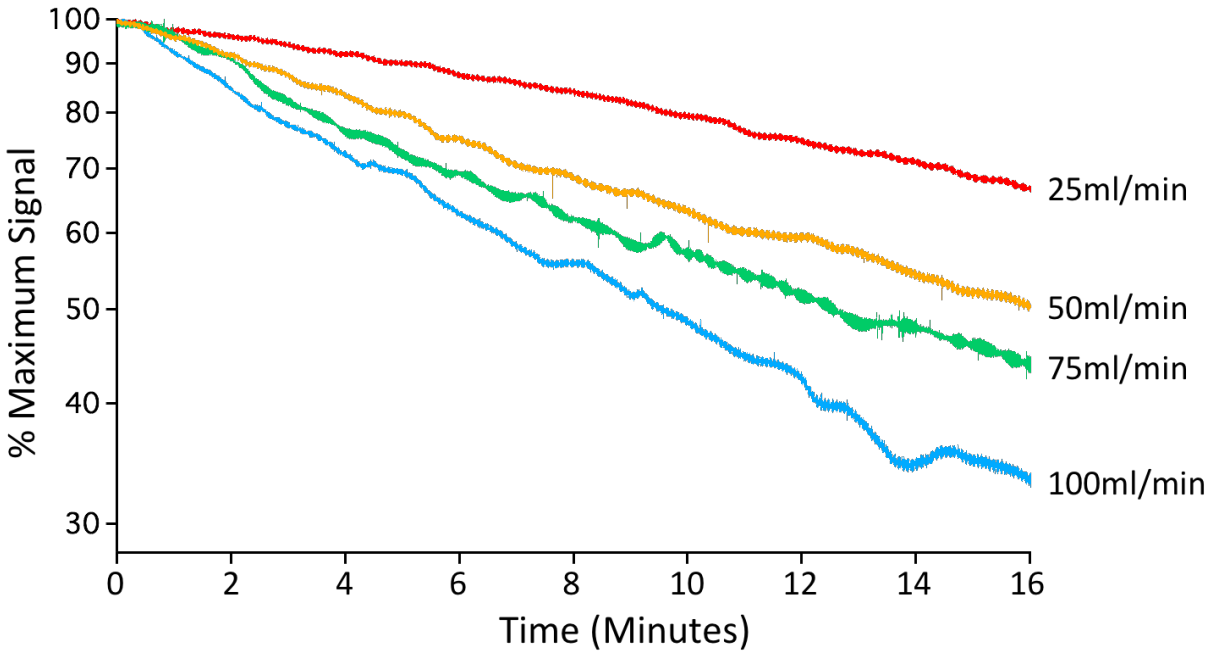


Figure 11.23: Results of dilution experiments to simulate different degrees of renal dysfunction, from CKD1 – CKD4, equivalent to creatinine clearance rates of 100ml/min – 25ml/min. Log-linear plot of raw signals normalised against the maximum observed for each experiment and time-aligned

The half lives of these curves are given in Table 11.6 on page 371. For all but the 25ml/min clearance experiment, these were read directly from the graph. The final value was obtained by extrapolation. Whilst these results do not directly correspond to the experimental design, they do follow an ordered sequence with some proportionality between the values obtained.

The results are notably more stable at lower dilution rates, further implicating dialysis recovery and mixing as sources of error.

Clearance Rate (ml/min)	Half-Life
100ml/min	9.5min
75ml/min	12.5min
50ml/min	16min
25ml/min	24min

Table 11.6: Half-lives calculated by dilution experiments to simulate different degrees of renal dysfunction, from CKD1 – CKD4, equivalent to creatinine clearance rates of 100ml/min – 25ml/min

Testing the System with a Blood-Perfused Porcine Kidney

The final experiment explored the function of this system in an isolated perfused kidney setup. To this end, I partnered with Dr. Bynvant Sandhu, a clinical researcher working at Hammersmith Hospital, one of the UK's major renal transplantation centres. Her work involved warm blood perfusion of porcine kidneys using an RM3 perfusion device[1]. An adult pig kidney was collected from a nearby licensed abbatoir and maintained in static cold storage for 4 hours. Following this, it was connected into an RM3 perfusion device which had been reconfigured with a heat exchanger and oxygenator for warm perfusion. The autologous blood collected for the reperfusion experiment was visibly hæmolysed and contained large amounts of thrombus which had to be filtered out prior to use.

After calibrating the sensor system against 100μM creatinine directly infused into the Y-junction and then via the microdialysis probe in an unstirred 100μM creatinine-T1 solution, I placed the probe tip deep into the stump of the renal vein to ensure good flow. Figure 11.24 shows the experimental setup in more detail. Data was then collected over the next hour of reperfusion until the probe membrane became damaged during repositioning and the experiment had to be abandoned.

[1] Waters Medical Systems LLC, Rochester, Minnesota, USA

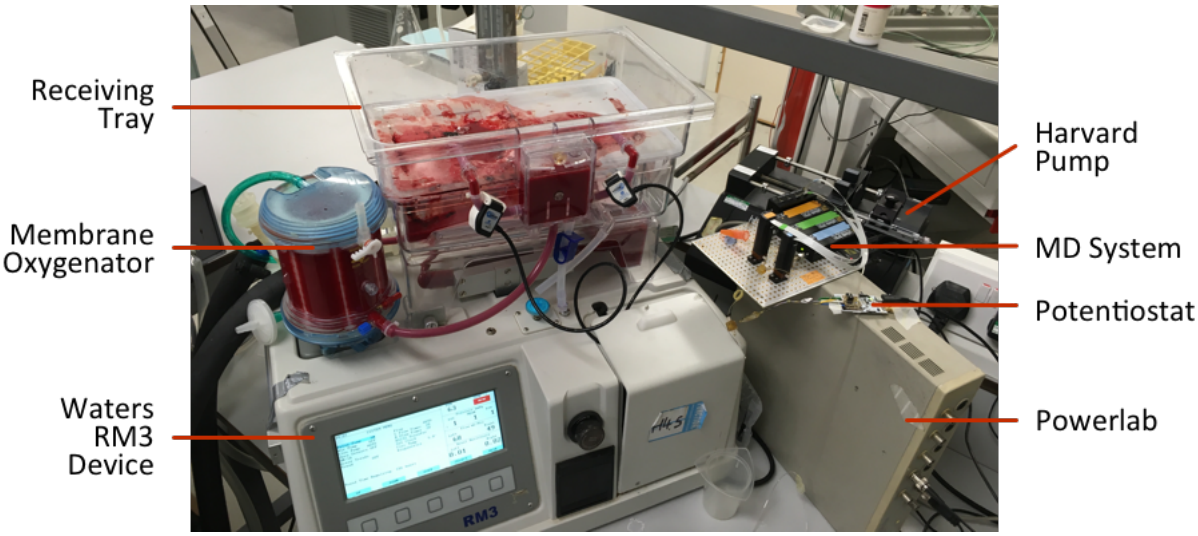


Figure 11.24: The Waters RM3 cold perfusion system configured for warm blood perfusion with an external membrane oxygenator and heat exchanger (out of frame). The microdialysis system has completed initial calibrations and is waiting for the kidney to arrive

Data analysis first required the use of a Savitsky-Golay smoothing filter[1] to remove the visible electrical spikes caused by the RM3's perfusion pump, as shown in Figure 11.25 below.

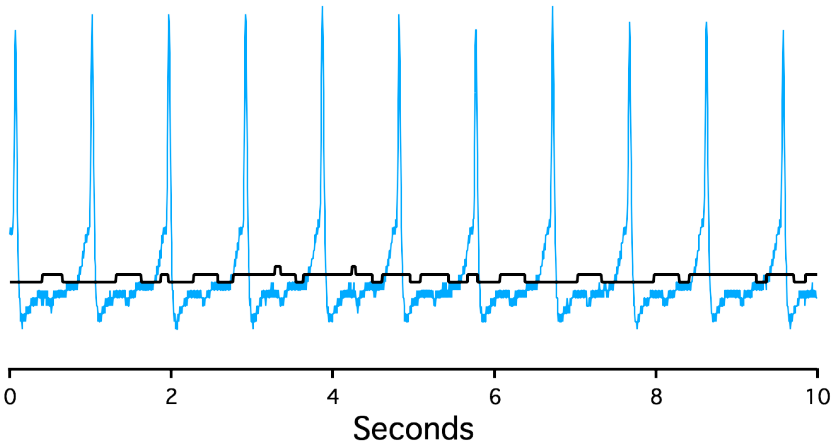


Figure 11.25: **Blue:** Raw signal from the microdialysis system showing the regular electrical spikes from the RM3's pump **Black:** Results of applying a Savitsky-Golay smoothing filter to the data

Figure 11.26 below shows an overview of the entire experiment, excluding the prior calibrations.

[1] 2nd order polynomial with a window of 513 samples

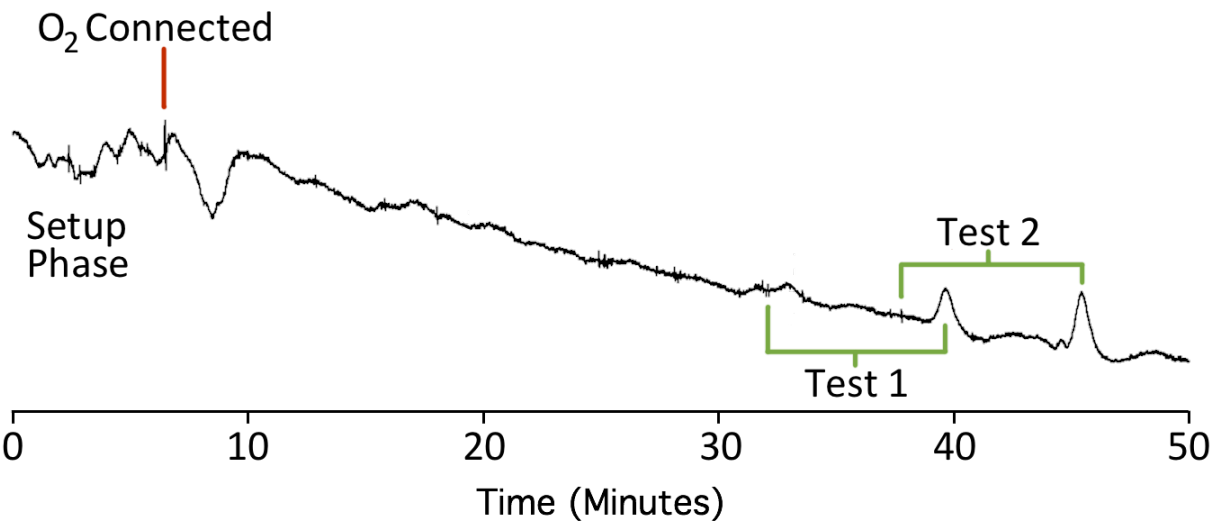


Figure 11.26: Results of the warm perfusion experiments showing an initial plateau phase followed by a steady decrease in signal magnitude following oxygenation, and the two subsequent creatinine tests

These results show an initial plateau during system setup and initial perfusion, equivalent to $\approx$300μM creatinine. This high system offset is probably due to a combination of the muscle damage from the slaughtering process, and extensive hæmolysis of the autologous blood, releasing creatine into the perfusate. When perfusion first began, the blood noticeably darkened as the kidney began consuming oxygen. Opening up the oxygen supply to the membrane oxygenator quickly returned the blood to a ruby red colour and caused a sudden decrease in the signal magnitude which soon returned to the high baseline. This may have in fact reflected a sudden oxidative burst from the ischæmic kidney consuming the oxygen required for the sarcosine oxidase to function normally, or a rapid change in pH which was detected by the sensor.

The kidney then appeared to be excreting detectable metabolites at a rate lying between that of the previous 100ml/min and 75ml/min creatinine clearance experiments (page 370), with a half-life of 652$\pm$3.5 seconds ($\approx$11min), with the caveat that the results may not be entirely equivalent. I then spiked the arterial reservoir of the RM3 system with two separate aliquots of 100μmoles of creatinine[1], producing the results seen in Figure 11.26. These curves had half-lives of 27 seconds and 18 seconds respectively, indicating that these results were more likely due to immediate dilution than clearance by the kidney.

Unfortunately the experiment had to be ended before the detectable metabolites in the system had been reduced to a low steady state. In the final reperfusion system as imagined, the perfusate would comprise washed erythrocytes in an isotonic crystalloid solution without any endogenous creatinine, thus allowing for pure clearance testing.

[1] 10mls$\times$10mM

11.4 Conclusions and Future Work

This part of the project has shown that a self-contained system based upon microdialysis sampling and amperometric testing of creatinine is able to achieve a limit of detection of 4.3μM and tested upper range of 500μM, matching or exceeding those reported in the literature (Table 11.3 on page 356). This performance was due to a series of improvements and optimisations I made to the potentiostat, microelectrode sensor array and the triple-enzyme system of Tsuchida and Yoda[149]. The process of electropolymerising mPD onto the working electrode also provided good protection against levels of interferents far in excess of those reported by other groups performing such testing.

In addition to the development of a real-time creatinine monitoring system[1] I have proposed and explored a novel way to monitor renal function without sensor calibration, thereby avoiding the need to compensate for any background noise or change in sensor offset, drift, or loss of sensitivity over time. I believe that this can be achieved by measuring the time constant[2] of the decay curve of creatinine excretion, and have demonstrated this experimentally in a closed-loop perfusion system containing a porcine kidney.

The economics of using this microfluidic system for real-time monitoring are also favourable. Despite the continuous wastage of the enzyme used in the analysis, a week of continuous monitoring would only consume 5ml of the 600:300:60 mixture. At current market prices for the three enzymes as of September 2016, this would amount to less than £50/week.

Future work would see the enzymes re-optimised in a buffer with higher pK_a such as HEPBS, the creation of a modular microdialysis sampling probe for in-line inclusion in a perfusion circuit, and an attempt to standardise the formation of a microelectrode array within a microchannel to provide a 'hot-pluggable' system for live creatinine monitoring.

The microfluidics could also be reduced in scale and internal volume, and employ a miniaturised serpentine microfluidic mixer instead of the hand-wound mixer shown in Figure 11.16 (page 363) so as to improve mixing and reduce the overall reaction time at even lower flow rates.

The system may also benefit from using droplet microfluidics (discussed on page 333 in Chapter 10) to allow the multiplexing of multiple enzyme reactions in parallel with a common sensor, whilst producing better mixing and less Taylor dispersion than is probably occurring in the present 3-minute delay loop. Reducing Taylor dispersion would have the added benefit of preserving transient or rapid changes in substrate concentration in the

[1] With a 3 minute delay for reaction time
[2] Or half-life

system, which are probably being lost at present.

With further time for development, the system would first undergo a period of *in vitro* refinement in the porcine kidney apparatus to improve the sensitivity and stability of the sensors and enzyme delivery pumps, and to create methods for compensating for these sources of error. This would be followed by a period of ambulatory *in vivo* testing in a small animal model, to learn more about live renal function which are not elucidated by infrequent spot testing of blood creatinine levels. This would determine the potential for translation into a clinical setting, where the final target would be the live monitoring of human renal function in an isolated perfusion apparatus for donor organs, and in an intensive-care setting where patients are at greatest risk of renal insults.

Overall, the three components of this Thesis – the Tesla-type pump, the live in-line multiparametric monitoring system, and the live renal function monitor – comprise a solid foundation bringing us closer to the goal of maintaining organs in optimal condition prior to transplantation, buying time in a setting where every second counts.

Bibliography

[1] Group TAPPK. Increasing Uptake of Home Dialysis to Benefit Patients and the NHS: A Report of the Findings of the 2013 Home Dialysis Summit. 2013;.

[2] Kerr M, Bray B, Medcalf J, O'donoghue DJ, Matthews B. Estimating the financial cost of chronic kidney disease to the NHS in England. Nephrology Dialysis Transplantation. 2012 August;27((Supple 3)):iii73–iii80.

[3] Organ Donation and Transplantation Activity Report 2015/16. NHS Blood and Transplant; May 2016.

[4] Papalois VE, Hakim NS, Najarian JS. Ch. 4. In: History of organ and cell transplantation. 1st ed. World Scientific; 2003. p. 76–99.

[5] Morris PJ. Transplantation—a medical miracle of the 20th century. New England Journal of Medicine. 2004;351(26):2678–2680.

[6] Hamilton D. A history of organ transplantation: ancient legends to modern practice. University of Pittsburgh Press; 2012.

[7] Moran M. Historical Concept of Brain Death Still Makes Waves Today. Neurology Today. 2005;5(7):11–13.

[8] Beecher HK, Adams R, Barger AC, Curran W, et al. A definition of irreversible coma: report of the Ad Hoc Committee of the Harvard Medical School to examine the definition of brain death. JAMA. 1968;205(6):337–40.

[9] Kanellis J. Justification for living donor kidney transplantation. Nephrology. 2010;15(s1):S72–S79.

[10] Summers DM, Johnson RJ, Allen J, Fuggle SV, Collett D, Watson CJ, et al. Analysis of factors that affect outcome after transplantation of kidneys donated after cardiac death in the UK: a cohort study. The Lancet. 2010;376(9749):1303–1311.

[11] Kootstra G, Daemen J, Oomen A. Categories of non-heart-beating donors. In: Transplantation proceedings. vol. 27; 1995. p. 2893.

[12] Sanchez-Fructuoso AI, Prats D, Torrente J, Perez-Contin MJ, Fernandez C, Alvarez J, et al. Renal transplantation from non-heart beating donors a promising alternative to enlarge the donor pool. Journal of the American Society of Nephrology. 2000;11(2):350–358.

[13] Dunne K, Doherty P. Donation after circulatory death. Continuing Education in Anaesthesia Critical Care and Pain. 2011;11(3):82–86.

[14] Patel R, Mickey MR, Terasaki PI, et al. Serotyping for homotransplantation. XVI. Analysis of kidney transplants from unrelated donors. The New England Journal of Medicine. 1968;279(10):501.

[15] Carrel A. The culture of whole organs: I. technique of the culture of the thyroid gland. The Journal of Experimental Medicine. 1937;65(4):515.

[16] Bickford R, Winton F. The influence of temperature on the isolated kidney of the dog. The Journal of Physiology. 1937;89(2):198–219.

[17] Calne R, Pegg D, Pryse-Davies J, Brown FL. Renal preservation by ice-cooling. British Medical Journal. 1963;2(5358):640–644.

[18] Crosley Jr AP, Castillo C, Rowe GG. The relationship of renal oxygen consumption to renal function and weight in individuals with normal and diseased kidneys. Journal of Clinical Investigation. 1961;40(5):836.

[19] Kurnik B, Weisberg LS, Kurnik PB. Renal and systemic oxygen consumption in patients with normal and abnormal renal function. Journal of the American Society of Nephrology. 1992;2(11):1617–1626.

[20] Belzer F, Southard J. The future of kidney preservation. Transplantation. 1980;30(3):161–164.

[21] Taylor MJ, Baicu SC. Current state of hypothermic machine perfusion preservation of organs: The clinical perspective. Cryobiology. 2010;60(3):S20–S35.

[22] Guibert EE, Petrenko AY, Balaban CL, Somov AY, Rodriguez JV, Fuller BJ. Organ preservation: current concepts and new strategies for the next decade. Transfusion Medicine and Hemotherapy. 2011;38(2):125–142.

[23] Mosbah I, Franco-Gou R, Abdennebi H, Hernandez R, Escolar G, Saidane D, et al.; Elsevier. Effects of polyethylene glycol and hydroxyethyl starch in University of Wisconsin preservation solution on human red blood cell aggregation and viscosity. Transplantation Proceedings. 2006;38(5):1229–1235.

[24] Badet L, Petruzzo P, Lefrancois N, McGregor B, Espa M, Berthillot C, et al.; Elsevier. Kidney

preservation with IGL-1 solution: a preliminary report. Transplantation Proceedings. 2005;37(1):308–311.

[25] Fuller BJ, Lee CY. Hypothermic perfusion preservation: The future of organ preservation revisited? Cryobiology. 2007;54(2):129–145.

[26] Hosgood SA, Nicholson ML. The role of perfluorocarbon in organ preservation. Transplantation. 2010;89(10):1169.

[27] Moers C, Pirenne J, Paul A, Ploeg RJ, Machine Preservation Trial Study Group. Machine perfusion or cold storage in deceased-donor kidney transplantation. New England Journal of Medicine. 2012 Feb;366(8):770–771.

[28] Brasile L, Stubenitsky B, Booster M, Green E, Haisch C, Kootstra G; Elsevier. Application of exsanguineous metabolic support to human kidneys. Transplantation proceedings. 2001;33(1-2):964–965.

[29] Brasile L, Stubenitsky BM, Booster MH, Arenada D, Haisch C, Kootstra G. Hypothermia–a limiting factor in using warm ischemically damaged kidneys. American Journal of Transplantation. 2001;1(4):316–320.

[30] Brasile L, Stubenitsky BM, Booster MH, Lindell S, Araneda D, Buck C, et al. Overcoming severe renal ischemia: the role of ex vivo warm perfusion. Transplantation. 2002;73(6):897–901.

[31] Brasile L, Stubenitsky BM, Haisch CE, Kon M, Kootstra G. Repair of damaged organs in vitro. American Journal of Transplantation. 2005;5(2):300–306.

[32] Brasile L, Glowacki P, Castracane J, Stubenitsky BM. Pretransplant kidney-specific treatment to eliminate the need for systemic immunosuppression. Transplantation. 2010;90(12):1294–1298.

[33] Hosgood SA, Barlow AD, Yates PJ, Snoeijs MG, van Heurn EL, Nicholson ML. A pilot study assessing the feasibility of a short period of normothermic preservation in an experimental model of non heart beating donor kidneys. Journal of Surgical Research. 2011;171(1):283–290.

[34] Hosgood SA, Nicholson ML. First in man renal transplantation after ex vivo normothermic perfusion. Transplantation. 2011;92(7):735–738.

[35] Hosgood S, Saeb-Parsy K, Hamed M, Nicholson M. Successful transplantation of human kidneys deemed untransplantable but resuscitated by ex-vivo normothermic machine perfusion. American Journal of Transplantation. 2016;.

[36] Hosgood SA, Heurn E, Nicholson ML. Normothermic machine perfusion of the kidney: better conditioning and repair? Transplant International. 2015;28(6):657–664.

[37] Hamaoui K, Smith R, Gowers S, Mastoridis S, Boutelle M, Darzi A, et al. Graft Preconditioning Using Localising Anticoagulant Fusion Proteins In Kidney Transplantation. Transplant International. 2013;26:67.

[38] Ganong WF. Renal Function & Micturition. In: Review of Medical Physiology. 22nd ed. Lange Medical Books/McGraw-Hill; 2005. p. 699–738.

[39] OpenStax, The Kidneys and Osmoregulatory Organs. OpenStax CNX.;. Available from: http://cnx.org/contents/3e2c416b-efb4-4386-a1ca-f18fb7b83ec5@8.

[40] OpenStax, Microscopic Anatomy of the Kidney. OpenStax CNX.;. Available from: http://cnx.org/contents/5c9bdf05-dd60-4769-a25a-e2958ff4c7a1@4.

[41] Huang Foen Chung JWNC, van Mastrigt R. Age and Volume Dependent Normal Frequency Volume Charts for Healthy Males. The Journal of Urology. 2009;182(1):210–214.

[42] Ganong WF. Circulating Body Fluids. In: Review of Medical Physiology. 22nd ed. Lange Medical Books/McGraw-Hill; 2005. p. 515–546.

[43] Ganong WF. Regulation of Extracellular Fluid Composition & Volume. In: Review of Medical Physiology. 22nd ed. Lange Medical Books/McGraw-Hill; 2005. p. 729–738.

[44] Wenger RH, Hoogewijs D. Regulated oxygen sensing by protein hydroxylation in renal erythropoietin-producing cells. American Journal of Physiology – Renal Physiology. 2010;298(6):F1287–F1296.

[45] Mount P, Power DA. Nitric oxide in the kidney: functions and regulation of synthesis. Acta Physiologica. 2006;187(4):433–446.

[46] Ferrante AM, Moscato U, Colacchio EC, Snider F. Results after elective open repair of pararenal abdominal aortic aneurysms. Journal of Vascular Surgery. 2016;63(6):1443–1450.

[47] Redfors B, Bragadottir G, Sellgren J, Swärd K, Ricksten SE. Effects of norepinephrine on renal perfusion, filtration and oxygenation in vasodilatory shock and acute kidney injury. Intensive Care Medicine. 2011;37(1):60–67.

[48] Monahan-Earley R, Dvorak A, Aird W. Evolutionary origins of the blood vascular system and

endothelium. Journal of Thrombosis and Haemostasis. 2013;11(s1):46–66.

[49] Collins JA, Rudenski A, Gibson J, Howard L, O'Driscoll R. Relating oxygen partial pressure, saturation and content: the haemoglobin–oxygen dissociation curve. Breathe. 2015;11(3):194.

[50] Roughton F, Severinghaus J. Accurate determination of 02 dissociation curve of human blood above 98.7% saturation with data on O2 solubility in unmodified human blood from 0 degrees to 37 degrees C. J Appl Physiol. 1973;35(6):861–869.

[51] Ramsing N, Gundersen J. Seawater and gases. Limnol Oceanogr. 2011;37:1307–1312.

[52] Brezis M, Rosen S. Hypoxia of the Renal Medulla — Its Implications for Disease. New England Journal of Medicine. 1995;332(10):647–655. PMID: 7845430. Available from: http://dx.doi.org/10.1056/NEJM199503093321006.

[53] Dagher PC. Apoptosis in ischemic renal injury: roles of GTP depletion and p53. Kidney International. 2004;66(2):506–509.

[54] Basile DP, Anderson MD, Sutton TA. Pathophysiology of acute kidney injury. Comprehensive Physiology. 2012;.

[55] Klahr S, Hamm LL, Hammerman MR, Mandel LJ. Renal Metabolism: Integrated Responses. Comprehensive Physiology. 2011;p. 2263–2333.

[56] Cargill WH, Hickam JB. The oxygen consumption of the normal and the diseased human kidney. Journal of Clinical Investigation. 1949;28(3):526.

[57] O'Connor PM. Renal oxygen delivery: Matching delivery to metabolic demand. Clinical and Experimental Pharmacology and Physiology. 2006;33(10):961–967.

[58] Weegman B, Kirchner V, Scott W, Avgoustiniatos E, Suszynski T, Ferrer-Fabrega J, et al. Continuous real-time viability assessment of kidneys based on oxygen consumption. In: Transplantation Proceedings. vol. 42. Elsevier; 2010. p. 2020–2023.

[59] Baverel G, Ferrier B, Martin M. Fuel selection by the kidney: adaptation to starvation. Proceedings of the Nutrition Society. 1995;54(01):197–212.

[60] Guder W, Schmolke M. Renal energy metabolism. In: Diuretics. Springer; 1995. p. 115–140.

[61] Bonventre JV, Vaidya VS, Schmouder R, Feig P, Dieterle F. Next-Generation Biomarkers for Detecting Kidney Toxicity. Nature Biotechnology. 2010;28(5):436–440.

[62] Soni SS, Ronco C, Katz N, Cruz DN. Early Diagnosis of Acute Kidney Injury: The Promise of Novel Biomarkers. Blood Purification. 2009;28:165–174.

[63] Kronenberg F. Emerging Risk Factors and Markers of Chronic Kidney Disease Progression. Nature Reviews Nephrology. 2009;5(12):677–689.

[64] Hall IE, Yarlagadda SG, Coca SG, Wang Z, Doshi M, Devarajan P, et al. IL-18 and Urinary NGAL Predict Dialysis and Graft Recovery after Kidney Transplantation. J Am Soc Nephrol. 2010;21(1):189–197. Available from: http://jasn.asnjournals.org/cgi/content/abstract/21/1/189.

[65] Haase-Fielitz A, Bellomo R, Devarajan P, Story D, Matalanis G, Dragun D, et al. Novel and Conventional Serum Biomarkers Predicting Acute Kidney Injury in Adult Cardiac Surgery-A Prospective Cohort Study. Critical Care Medicine. 2009;37(2):553.

[66] Devarajan P. Neutrophil Gelatinase-Associated Lipocalin: A Promising Biomarker for Human Acute Kidney Injury. Biomarkers. 2010;4(2):265–280.

[67] Dent C, Ma Q, Dastrala S, Bennett M, Mitsnefes M, Barasch J, et al. Plasma Neutrophil Gelatinase-Associated Lipocalin Predicts Acute Kidney Injury, Morbidity and Mortality after Pediatric Cardiac Surgery: A Prospective Uncontrolled Cohort Study. Critical Care. 2007;11(6):R127. Available from: http://ccforum.com/content/11/6/R127.

[68] Bagshaw S, Bennett M, Haase M, Haase-Fielitz A, Egi M, Morimatsu H, et al. Plasma and Urine Neutrophil Gelatinase-Associated Lipocalin in Septic Versus Non-Septic Acute Kidney Injury in Critical Illness. Intensive Care Medicine. 2010;36:452–461. 10.1007/s00134-009-1724-9. Available from: http://dx.doi.org/10.1007/s00134-009-1724-9.

[69] Rehberg PB. Studies on kidney function: the rate of filtration and reabsorption in the human kidney. Biochemical Journal. 1926;20(3):447.

[70] Hsu Cy, Bansal N. Measured GFR as "gold standard"—all that glitters is not gold? Clinical Journal of the American Society of Nephrology. 2011;6(8):1813–1814.

[71] Botev R, Mallié JP, Couchoud C, Schück O, Fauvel JP, Wetzels JF, et al. Estimating glomerular filtration rate: Cockcroft–Gault and Modification of Diet in Renal Disease formulas compared to renal inulin clearance. Clinical Journal of the American Society of Nephrology. 2009;4(5):899–906.

[72] Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. Nephron.

1976;16(1):31–41.

[73] Klahr S, Levey AS, Beck GJ, Caggiula AW, Hunsicker L, Kusek JW, et al. The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease. New England Journal of Medicine. 1994;330(13):877–884.

[74] Levey AS, Coresh J, Greene T, Marsh J, Stevens LA, Kusek JW, et al. Expressing the Modification of Diet in Renal Disease Study equation for estimating glomerular filtration rate with standardized serum creatinine values. Clinical Chemistry. 2007;53(4):766–772.

[75] Westerhof N. Cardiac Work and Efficiency. Cardiovascular Research. 2000;48(1):4–7.

[76] Ganong WF. The Heart as a Pump. In: Review of Medical Physiology. 22nd ed. Lange Medical Books/McGraw-Hill; 2005. p. 565–576.

[77] Tayama E, Raskin SA, Nosé Y. Chapter 3: Blood Pumps. In: Gravlee GP, Davis RF, Kurusz M, Utley JR, editors. Cardiopulmonary Bypass: Principles and Practice. 3rd ed. Lippincott Williams & Wilkins; 2000. p. 35–46.

[78] Sousa PC, Pinho FT, Alves MA, Oliveira MS. A review of hemorheology: Measuring techniques and recent advances. Korea-Australia Rheology Journal. 2016;28(1):1–22.

[79] Hosseini SM, Feng JJ. A particle-based model for the transport of erythrocytes in capillaries. Chemical Engineering Science. 2009;64(22):4488–4497.

[80] Rother RP, Bell L, Hillmen P, Gladwin MT. The clinical sequelae of intravascular hemolysis and extracellular plasma hemoglobin: a novel mechanism of human disease. JAMA. 2005;293(13):1653–1662.

[81] Windsant ICV, Hanssen SJ, Buurman WA, Jacobs MJ. Cardiovascular surgery and organ damage: time to reconsider the role of hemolysis. The Journal of Thoracic and Cardiovascular Surgery. 2011;142(1):1–11.

[82] Vercaemst L. Hemolysis in cardiac surgery patients undergoing cardiopulmonary bypass: a review in search of a treatment algorithm. The Journal of Extra-Corporeal Technology. 2008;40(4):257–267.

[83] Mamikonian LS, Mamo LB, Smith PB, Koo J, Lodge AJ, Turi JL. Cardiopulmonary Bypass is Associated with Hemolysis and Acute Kidney Injury in Neonates, Infants and Children. Pediatric Critical Care Medicine. 2014;15(3):e111–e119.

[84] Hathcock JJ. Flow effects on coagulation and thrombosis. Arteriosclerosis, Thrombosis, and Vascular Biology. 2006;26(8):1729–1737.

[85] Smith RL, Blick EF, Coalson J, Stein PD. Thrombus production by turbulence. Journal of Applied Physiology. 1972;32(2):261–264.

[86] Nesbitt WS, Mangin P, Salem HH, Jackson SP. The impact of blood rheology on the molecular and cellular events underlying arterial thrombosis. Journal of Molecular Medicine. 2006;84(12):989–995.

[87] Jones SA. A relationship between Reynolds stresses and viscous dissipation: implications to red cell damage. Annals of Biomedical Engineering. 1995;23(1):21–28.

[88] Sakariassen KS, Orning L, Turitto VT. The impact of blood shear rate on arterial thrombus formation. Future Science OA. 2015;1(4).

[89] Giersiepen M, Wurzinger L, Opitz R, Reul H. Estimation of shear stress-related blood damage in heart valve prostheses–in vitro comparison of 25 aortic valves. The International Journal of Artificial Organs. 1990;13(5):300–306.

[90] Arora D, Behr M, Coronado O, Pasquali M. Estimation of hemolysis in centrifugal blood pumps using morphology tensor approach. In: Proceedings of the Third MIT Conference on Computational Fluid and Solid Dynamics. Massachussetts: Elsevier, Cambridge; 2005. p. 578–82.

[91] Ewbank T. A descriptive and historical account of hydraulic and other machines for raising water, ancient and modern: with observations on various subjects connected with the mechanic arts: including the progressive development of the steam engine, etc. in five books. Derby & Jackson; 1857.

[92] A Brief History of Pumps. World Pumps. 2009;508:30–37.

[93] Yannopoulos SI, Lyberatos G, Theodossiou N, Li W, Valipour M, Tamburrino A, et al. Evolution of water lifting devices (pumps) over the centuries worldwide. Water. 2015;7(9):5031–5060.

[94] Black NH, Davis HN. XII. In: Practical physics for secondary schools. Fundamental principles and applications to daily life. Macmillan and Company; 1913. p. 219.

[95] Boettcher W, Merkle F, Weitkemper HH. History of extracorporeal circulation: the invention and modification of blood pumps. The Journal of Extra-Corporeal Technology. 2003;35(3):184–191.

[96] DeBakey ME. Left ventricular bypass pump for cardiac assistance: clinical experience. The American Journal of Cardiology. 1971;27(1):3–11.

[97] Copeland JG, Smith RG, Arabia FA, Nolan PE, Sethi GK, Tsau PH, et al. Cardiac replacement with a total artificial heart as a bridge to transplantation. New England Journal of Medicine. 2004;351(9):859–867.

[98] Rodriguez LE, Suarez EE, Loebe M, Bruckner BA. Ventricular assist devices (VAD) therapy: new technology, new hope? Methodist DeBakey Cardiovascular Journal. 2013;9(1):32–37.

[99] Kijima T, Oshiyama H, Horiuchi K, Nogawa A, Hamasaki H, Amano N, et al. A straight path centrifugal blood pump concept in the Capiox centrifugal pump. Artificial Organs. 1993;17(7):593–598.

[100] Kawahito K, Nosé Y. Hemolysis in different centrifugal pumps. Artificial Organs. 1997;21(4):323–326.

[101] Noon GP, Ball JW, Papaconstantinou HT. Clinical experience with BioMedicus centrifugal ventricular support in 172 patients. Artificial Organs. 1995;19(7):756–760.

[102] Ganong WF. Cardiovascular Regulatory Mechanisms. In: Review of Medical Physiology. 22nd ed. Lange Medical Books/McGraw-Hill; 2005. p. 597–602.

[103] Cheng A, Williamitis CA, Slaughter MS. Comparison of continuous-flow and pulsatile-flow left ventricular assist devices: is there an advantage to pulsatility? Annals of Cardiothoracic Surgery. 2014;3(6):573–581.

[104] Cleland J, McDonagh T, Mitchell P. National Heart Failure Audit, April 2012-March 2013. NICOR (National Institute for Cardiovascular Outcomes Research) and The British Society for Heart Failure (BSH). 2013;.

[105] Tesla N. Turbine. Google Patents; 1913. US Patent 1,061,206.

[106] Hasinger S, Kehrt L. Investigation of a shear-force pump. Journal of Engineering for Gas Turbines and Power. 1963;85(3):201–206.

[107] Izraelev V, Weiss WJ, Fritz B, Newswanger RK, Paterson EG, Snyder A, et al. A passively-suspended Tesla pump left ventricular assist device. ASAIO Journal. 2009;55(6):556.

[108] Medvitz RB, Boger DA, Izraelev V, Rosenberg G, Paterson EG. Computational fluid dynamics design and analysis of a passively suspended Tesla pump left ventricular assist device. Artificial Organs. 2011;35(5):522–533.

[109] Brennen CE. 4. In: Hydrodynamics of Pumps (A Concepts ETI Book). Oxford University Press; 1995. p. 74.

[110] Sochi T. Non-Newtonian Rheology in Blood Circulation. ArXiv e-prints. 2013 June;(1306.2067). Available from: http://adsabs.harvard.edu/abs/2013arXiv1306.2067S.

[111] Fedosov DA, Caswell B, Popel AS, Karniadakis GE. Blood Flow and Cell-Free Layer in Microvessels. Microcirculation. 2010;17(8):615–628.

[112] Miller GE, Etter BD, Dorsi JM. A multiple disk centrifugal pump as a blood flow device. IEEE Transactions on Biomedical Engineering. 1990;37(2):157–163.

[113] Miller GE, Fink R. Analysis of optimal design configurations for a multiple disk centrifugal blood pump. Artificial Organs. 1999;23(6):559–565.

[114] Jhun CS, Newswanger R, Cysyk J, Lukic B, Weiss W, Rosenberg G. Tesla-Based Blood Pump and Its Applications. Journal of Medical Devices. 2013;7(4):040917.

[115] Breiter MC, Pohlhausen K. Laminar flow between two parallel rotating disks. Office of Aerospace Research, Arlington, VA; 1962. ARL-62-318.

[116] Matsch L, Rice W. An asymptotic solution for laminar flow of an incompressible fluid between rotating disks. Journal of Applied Mechanics. 1968;35(1):155–159.

[117] Lawn MJ Jr, Rice W. Calculated Design Data for the Multiple-Disk Turbine Using Incompressible Fluid. Journal of Fluids Engineering. 1974 09;96(3):252–258. Available from: http://dx.doi.org/10.1115/1.3447148.

[118] Oliveira MD, Páscoa JC. Analytical and experimental modeling of a viscous disc pump for MEMS applications. In: III Conferência Nacional em Mecânica de Fluidos, Termodinâmica e Energia (MEFTE – Bragança); 2009. .

[119] Kameneva MV, Repko BM, Krasik EF, Perricelli BC, Borovetz HS. Polyethylene glycol additives reduce hemolysis in red blood cell suspensions exposed to mechanical stress. ASAIO Journal. 2003;49(5):537–542.

[120] Zingg W, Sulev JC, Morgan CD. Relationship between viscosity and hematocrit in blood of normal persons and burn patients. Canadian Journal of Physiology and Pharmacology. 1970;48(3):202–205.

[121] Grover WH, Bryan AK, Diez-Silva M, Suresh S, Higgins JM, Manalis SR. Measuring single-cell

density. PNAS. 2011;108(27):10992–10996.

[122] Satyapal K, Rambiritch V, Pillai G. Morphometric analysis of the renal veins. The Anatomical Record. 1995;241(2):268–272.

[123] Mayevsky A, Sonn J, Luger-Hamer M, Nakache R. Real-Time assessment of organ vitality during the transplantation procedure. Transplantation Reviews. 2003;17(2):96 – 116.

[124] Potkay JA. Long term, implantable blood pressure monitoring systems. Biomedical Microdevices. 2008;10(3):379–392.

[125] bq2407x 1.5-A USB-Friendly Li-Ion Battery Charger and Power-Path Management IC; 2015.

[126] Zhang HJ. Basic Concepts of Linear Regulator and Switching Mode Power Supplies; 2013. Application Note 140, Analog Devices.

[127] Low Power Multichannel Thermocouple Measurement System with Cold Junction Compensation; 2014. Circuit Note 0354, Analog Devices.

[128] Company BC. Reference Tables for Thermocouple and RTD Sensors; 1996. Product Information Note 1110-ML-005-0-01.

[129] Hoskins PR. Physical properties of tissues relevant to arterial ultrasound imaging and blood velocity measurement. Ultrasound in Medicine & Biology. 2007;33(10):1527–1539.

[130] Constant Temperature Anemometer (CTA) Version 4; 2013. Application Note 001, IST AG.

[131] Monteleone M, Yeung H, Smith R. A review of Ancient Roman water supply exploring techniques of pressure reduction. Water Science and Technology: Water Supply. 2007;7(1):113–120.

[132] Roriz P, Frazão O, Lobo-Ribeiro AB, Santos JL, Simões JA. Review of fiber-optic pressure sensors for biomedical and biomechanical applications. Journal of Biomedical Optics. 2013;18(5):050903–050903.

[133] Chen LT, Chang JS, Hsu CY, Cheng WH. Fabrication and performance of MEMS-based pressure sensor packages using patterned ultra-thick photoresists. Sensors. 2009;9(8):6200–6218.

[134] Colvin AE, Phillips TE, Miragliotta JA, Givens RB, Bargeron CB. A novel solid-state oxygen sensor. Johns Hopkins Apl Technical Digest. 1996;17(4):377–385.

[135] McDonagh C, Kolle C, McEvoy A, Dowling D, Cafolla A, Cullen S, et al. Phase fluorometric dissolved oxygen sensor. Sensors and Actuators B: Chemical. 2001;74(1):124–130.

[136] Matcher S, Elwell C, Cooper C, Cope M, Delpy D. Performance comparison of several published tissue near-infrared spectroscopy algorithms. Analytical Biochemistry. 1995;227(1):54–68.

[137] Nitzan M, Romem A, Koppel R. Pulse oximetry: fundamentals and technology update. Medical Devices (Auckland, NZ). 2014;7:231.

[138] Texas Instruments LMP9007x/LMP90080 Sensor AFE System: Multi-Channel, Low-Power 16-Bit Sensor AFE with True Continuous Background Calibration; 2015. Available from: http://www.ti.com/product/LMP90080.

[139] Texas Instruments AFE4490 Integrated Analog Front-End for Pulse Oximeters; 2014. Available from: http://www.ti.com/product/AFE4490.

[140] Grosenbaugh DA, Alben JO, Muir WW. Absorbance Spectra of Inter-Species Hemoglobins in the Visible and Near Infrared Regions. Journal of Veterinary Emergency and Critical Care. 1997;7(1):36–42.

[141] Michel HO. A study of sulfhemoglobin. Journal of Biological Chemistry. 1938;126(1):323–348.

[142] Randviir EP, Banks CE. Analytical methods for quantifying creatinine within biological media. Sensors and Actuators B: Chemical. 2013;183:239–252.

[143] Delanghe JR, Speeckaert MM. Creatinine determination according to Jaffe—what does it stand for? NDT Plus. 2011;4(2):83–86.

[144] Myers GL, Miller WG, Coresh J, Fleming J, Greenberg N, Greene T, et al. Recommendations for improving serum creatinine measurement: a report from the Laboratory Working Group of the National Kidney Disease Education Program. Clinical Chemistry. 2006;52(1):5–18.

[145] Meites S. Otto Folin's medical legacy. Clinical Chemistry. 1985;31(8):1402–1404.

[146] Panteghini M. Enzymatic assays for creatinine: time for action. Scandinavian Journal of Clinical and Laboratory Investigation. 2008;68(sup241):84–88.

[147] Piéroni L, Delanaye P, Boutten A, Bargnoux AS, Rozet E, Delatour V, et al. A multicentric evaluation of IDMS-traceable creatinine enzymatic assays. Clinica Chimica Acta. 2011;412(23):2070–2075.

[148] Guilbault GG, Coulet PR. Creatinine-selective enzyme electrodes. Analytica Chimica Acta. 1983;152:223–228.

[149] Tsuchida T, Yoda K. Multi-enzyme membrane electrodes for determination of creatinine and creatine in serum. Clinical Chemistry. 1983;29(1):51–55.

[150] Feuerstein D, Manning A, Hashemi P, Bhatia R, Fabricius M, Tolias C, et al. Dynamic metabolic response to multiple spreading depolarizations in patients with acute brain injury: an online microdialysis study. Journal of Cerebral Blood Flow and Metabolism. 2010;30(7):1343–1355.

[151] Rogers ML, Brennan PA, Leong CL, Gowers SAN, Aldridge T, Mellor TK, et al. Online rapid sampling microdialysis (rsMD) using enzyme-based electroanalysis for dynamic detection of ischaemia during free flap reconstructive surgery. Anal Bioanal Chem. 2013 Feb;405(11):3881–3888. Available from: http://dx.doi.org/10.1007/s00216-013-6770-z.

[152] Rogers ML, Feuerstein D, Leong CL, Takagaki M, Niu X, Graf R, et al. Continuous Online Microdialysis Using Microfluidic Sensors: Dynamic Neurometabolic Changes during Spreading Depolarization. ACS Chem Neurosci. 2013 May;4(5):799–807. Available from: http://dx.doi.org/10.1021/cn400047x.

[153] Collison M, Meyerhoff M. Continuous-flow enzymatic determination of creatinine with improved on-line removal of endogenous ammonia. Analytica Chimica Acta. 1987;200:61–72.

[154] Radoi A, Compagnone D. Recent advances in NADH electrochemical sensing design. Bioelectrochemistry. 2009;76(1):126–134.

[155] Rui CS, Hiroaki IO, Sonomoto K, Kato Y. A Multifunctional Flow-injection Biosensor for the Simultaneous Determination of Ammonia, Creatinine, and Urea. Annals of the New York Academy of Sciences. 1995;750(1):30–38.

[156] Suzuki K, Siswanta D, Otsuka T, Amano T, Ikeda T, Hisamoto H, et al. Design and Synthesis of a More Highly Selective Ammonium Ionophore Than Nonactin and Its Application as an Ion-Sensing Component for an Ion-Selective Electrode. Analytical Chemistry. 2000 may;72(10):2200–2205. Available from: http://dx.doi.org/10.1021/ac9911241.

[157] Radomska A, Bodenszac E, Głąb S, Koncki R. Creatinine biosensor based on ammonium ion selective electrode and its application in flow-injection analysis. Talanta. 2004;64(3):603–608.

[158] Lähdesmäki I, Lewenstam A, Ivaska A. A polypyrrole-based amperometric ammonia sensor. Talanta. 1996 Jan;43(1):125–134. Available from: http://dx.doi.org/10.1016/0039-9140(95)01713-5.

[159] Timmer B, Olthuis W, Van Den Berg A. Ammonia sensors and their applications—a review. Sensors and Actuators B: Chemical. 2005;107(2):666–677.

[160] Xu JC, Hunter GW, Lukco D, Liu CC, Ward BJ. Novel carbon dioxide microsensor based on tin oxide nanomaterial doped with copper oxide. IEEE Sensors Journal. 2009;9(3):235–236.

[161] Suzuki H, Arakawa H, Karube I. Performance characteristics of a urea microsensor employing a micromachined carbon dioxide electrode. Electroanalysis. 2000;12(16):1327–1333.

[162] Halliwell B, Clement MV, Long LH. Hydrogen peroxide in the human body. FEBS letters. 2000;486(1):10–13.

[163] Watson CJ, Venton BJ, Kennedy RT. In vivo measurements of neurotransmitters by microdialysis sampling. Analytical Chemistry. 2006;78(5):1391–1399.

[164] Ungerstedt U, Pycock C. Functional correlates of dopamine neurotransmission. Bulletin der Schweizerischen Akademie der Medizinischen Wissenschaften. 1974;30(1-3):44–55.

[165] Whitesides GM. The origins and the future of microfluidics. Nature. 2006;442(7101):368–373.

[166] Rogers M, Leong C, Niu X, de Mello A, Parker KH, Boutelle MG. Optimisation of a microfluidic analysis chamber for the placement of microelectrodes. Phys Chem Chem Phys. 2011;13(12):5298. Available from: http://dx.doi.org/10.1039/C0CP02810J.

[167] Wang C. Development of clinical instruments for traumatic brain injury patients: design, realisation, and performance. Imperial College London; 2015.

[168] Patel BA, Rogers M, Wieder T, O'Hare D, Boutelle MG. ATP microelectrode biosensor for stable long-term in vitro monitoring from gastrointestinal tissue. Biosensors and Bioelectronics. 2011;26(6):2890–2896.

[169] Hall SB, Khudaish EA, Hart AL. Electrochemical oxidation of hydrogen peroxide at platinum electrodes. Part 1. An adsorption-controlled mechanism. Electrochimica Acta. 1998;43(5):579–588.

[170] Bartlett P, Pratt K. A study of the kinetics of the reaction between ferrocene monocarboxylic acid and glucose oxidase using the rotating-disc electrode. Journal of Electroanalytical Chemistry. 1995;397(1):53–60.

[171] Henzler T, Steudle E. Transport and metabolic degradation of hydrogen peroxide in Chara corallina: model calculations and measurements with the pressure probe suggest transport of H2O2 across water

channels. Journal of Experimental Botany. 2000;51(353):2053–2066.

[172] Neuman MR. Biopotential Electrodes. In: Bronzino JD, editor. The Biomedical Engineering Handbook: Second Edition. 48. CRC Press LLC, Boca Raton; 2000. .

[173] Liu H, Ying T, Sun K, Li H, Qi D. Reagentless amperometric biosensors highly sensitive to hydrogen peroxide, glucose and lactose based on N-methyl phenazine methosulfate incorporated in a Nafion film as an electron transfer mediator between horseradish peroxidase and an electrode. Analytica Chimica Acta. 1997;344(3):187–199.

[174] Guerrieri A, De Benedetto G, Palmisano F, Zambonin P. Electrosynthesized non-conducting polymers as permselective membranes in amperometric enzyme electrodes: a glucose biosensor based on a co-crosslinked glucose oxidase/overoxidized polypyrrole bilayer. Biosensors and Bioelectronics. 1998;13(1):103–112.

[175] Nakabayashi Y, Wakuda M, Imai H. Amperometric Glucose Sensors Fabricated by Electrochemical Polymerization of Phenols on Carbon Paste Electrodes Containing Ferrocene as an Electron Transfer Mediator. Analytical Sciences. 1998;14(6):1069–1076.

[176] Mitchell KM. Acetylcholine and Choline Amperometric Enzyme Sensors Characterized in Vitro and in Vivo. Analytical Chemistry. 2004 Feb;76(4):1098–1106. Available from: http://dx.doi.org/10.1021/ac034757v.

[177] Instruments T. LMP7721 3-Femtoampere Input Bias Current Precision Amplifier; 2014. Available from: http://www.ti.com/product/LMP7721.

[178] Weber SG. Signal-to-noise ratio in microelectrode-array-based electrochemical detectors. Analytical Chemistry. 1989;61(4):295–302.

[179] Fletcher S, Horne MD. Random assemblies of microelectrodes (RAM™ electrodes) for electrochemical studies. Electrochemistry Communications. 1999 oct;1(10):502–512. Available from: http://dx.doi.org/10.1016/S1388-2481(99)00100-9.

[180] Tomčík P. Microelectrode arrays with overlapped diffusion layers as electroanalytical detectors: theory and basic applications. Sensors. 2013;13(10):13659–13684.

[181] Khue NV, Wolff CM, Seris JL, Schwing JP. Immobilized enzyme electrode for creatinine determination in serum. Analytical Chemistry. 1991;63(6):611–614.

[182] Weber J, Van Zanten A. Interferences in current methods for measurements of creatinine. Clinical Chemistry. 1991;37(5):695–700.

[183] Sakslund H, Hammerich O. Effects of pH, temperature and reaction products on the performance of an immobilized creatininase-creatinase-sarcosine oxidase enzyme system for creatinine determination. Analytica Chimica Acta. 1992;268(2):331–345.

[184] Yamato H, Ohwa M, Wernet W. A polypyrrole/three-enzyme electrode for creatinine detection. Analytical Chemistry. 1995;67(17):2776–2780.

[185] Schneider J, Gründig B, Renneberg R, Cammann K, Madaras M, Buck R, et al. Hydrogel matrix for three enzyme entrapment in creatine/creatinine amperometric biosensing. Analytica Chimica Acta. 1996;325(3):161–167.

[186] Mădăraş MB, Popescu IC, Ufer S, Buck RP. Microfabricated amperometric creatine and creatinine biosensors. Analytica Chimica Acta. 1996;319(3):335–345.

[187] Mădăraş MB, Buck RP. Miniaturized biosensors employing electropolymerized permselective films and their use for creatinine assays in human serum. Analytical Chemistry. 1996;68(21):3832–3839.

[188] Khan G, Wernet W. A highly sensitive amperometric creatinine sensor. Analytica Chimica Acta. 1997;351(1):151–158.

[189] Kim EJ, Haruyama T, Yanagida Y, Kobatake E, Aizawa M. Disposable creatinine sensor based on thick-film hydrogen peroxide electrode system. Analytica Chimica Acta. 1999;394(2):225–231.

[190] Shin JH, Choi YS, Lee HJ, Choi SH, Ha J, Yoon IJ, et al. A planar amperometric creatinine biosensor employing an insoluble oxidizing agent for removing redox-active interferences. Analytical Chemistry. 2001;73(24):5965–5971.

[191] Tombach B, Schneider J, Matzkies F, Schaefer RM, Chemnitius GC. Amperometric creatinine biosensor for hemodialysis patients. Clinica Chimica Acta. 2001;312(1):129–134.

[192] Erlenkötter A, Fobker M, Chemnitius GC. Biosensors and flow-through system for the determination of creatinine in hemodialysate. Analytical and Bioanalytical Chemistry. 2002;372(2):284–292.

[193] Stefan RI, Bokretsion RG, van Staden JF, Aboul-Enein HY. Simultaneous determination of creatine and creatinine using amperometric biosensors. Talanta. 2003;60(6):1223–1228.

[194] Hsiue GH, Lu PL, Chen JC. Multienzyme-immobilized modified polypropylene membrane for an

amperometric creatinine biosensor. Journal of Applied Polymer Science. 2004;92(5):3126–3134.
[195] Berberich JA, Yang LW, Madura J, Bahar I, Russell AJ. A stable three-enzyme creatinine biosensor. 1. Impact of structure, function and environment on PEGylated and immobilized sarcosine oxidase. Acta Biomaterialia. 2005;1(2):173–181.
[196] Berberich JA, Yang LW, Bahar I, Russell AJ. A stable three enzyme creatinine biosensor. 2. Analysis of the impact of silver ions on creatine amidinohydrolase. Acta Biomaterialia. 2005;1(2):183–191.
[197] Berberich JA, Chan A, Boden M, Russell AJ. A stable three-enzyme creatinine biosensor. 3. Immobilization of creatinine amidohydrolase and sensor development. Acta Biomaterialia. 2005;1(2):193–199.
[198] Staden RISv, Bokretsion RG. Simultaneous Determination of Creatine and Creatinine using Monocrystalline Diamond Paste–Based Amperometric Biosensors. Analytical Letters. 2006;39(11):2227–2233.
[199] Nieh CH, Tsujimura S, Shirai O, Kano K. Amperometric biosensor based on reductive H_2O_2 detection using pentacyanoferrate-bound polymer for creatinine determination. Analytica Chimica Acta. 2013;767:128–133.
[200] Serafín V, Hernández P, Agüí L, Yáñez-Sedeño P, Pingarrón J. Electrochemical biosensor for creatinine based on the immobilization of creatininase, creatinase and sarcosine oxidase onto a ferrocene/horseradish peroxidase/gold nanoparticles/multi-walled carbon nanotubes/Teflon composite electrode. Electrochimica Acta. 2013;97:175–183.
[201] Yoshimoto T, Tanaka N, Kanada N, Inoue T, Nakajima Y, Haratake M, et al. Crystal structures of creatininase reveal the substrate binding site and provide an insight into the catalytic mechanism. Journal of Molecular Biology. 2004;337(2):399–416.
[202] Hall SB, Khudaish EA, Hart AL. Electrochemical oxidation of hydrogen peroxide at platinum electrodes. Part IV: phosphate buffer dependence. Electrochimica Acta. 1999;44(25):4573–4582.
[203] Ferreira CM, Pinto IS, Soares EV, Soares HM. (Un)suitability of the use of pH buffers in biological, biochemical and environmental studies and their interaction with metal ions–a review. RSC Advances. 2015;5(39):30989–31003.
[204] Judson G, Frauenfelder H, Mooney G. Biochemical Changes in Thoroughbred Racehorses Following Submaximal and Maximal Exercise. In: Equine Exercise Physiology. Granta Editions Cambridge; 1983. p. 408–415.
[205] Fang J, Alderman MH. Serum uric acid and cardiovascular mortality: the NHANES I epidemiologic follow-up study, 1971-1992. JAMA. 2000;283(18):2404–2410.
[206] Allen RH, Stabler SP, Lindenbaum J. Serum betaine, N, N-dimethylglycine and N-methylglycine levels in patients with cobalamin and folate deficiency and related inborn errors of metabolism. Metabolism. 1993 Nov;42(11):1448–1460. Available from: http://dx.doi.org/10.1016/0026-0495(93)90198-W.
[207] Schedel JM, Tanaka H, Kiyonaga A, Shindo M, Schutz Y. Acute creatine ingestion in human: consequences on serum creatine and creatinine concentrations. Life Sciences. 1999;65(23):2463–2470.
[208] Verhelst J, Berwaerts J, Marescau B, Abs R, Neels H, Mahler C, et al. Serum creatine, creatinine, and other guanidino compounds in patients with thyroid dysfunction. Metabolism. 1997;46(9):1063–1067.
[209] Albery WJ, Haggett BG, Svanberg LR. The development of electrochemical sensors. In: Advanced Agricultural Instrumentation. Springer; 1986. p. 349–392.