The Art of Oceanographic Instrumentation

Albert J. Williams 3rd

Contents

Introduction

Outline

Observation, models, understanding
Oceanography is an observational science
Instrumentation aids observation in all disciplines

Understanding is a powerful aid to prediction. Models lead to understanding, and observations ground models. Instruments and measurement techniques extend direct observation to the benefit of understanding and prediction. In oceanography, much is still to be learned through observation and the limit to observation is often instrumentation. Develop skill in utilization of oceanographic instrumentation and learn to specify or invent new instrumentation and you will aid observation. Better techniques of measurement lead to greater and more accurate understanding of the natural system. The short term benefit of such understanding is the ability to predict response of the ocean. In the long term, understanding is a general guard against blunders in our own behavior as a society. And understanding can lead to opportunities for human use and individual appreciation to society's benefit.

As a relative newcomer to scientific disciplines, oceanography is still in an exploratory stage where there is much to be learned by observation. Meteorology is only slightly older, and geology slightly older still, yet all three are as much limited by facts as by models. In physical oceanography the dominant questions are, "Where does the water go?" and, "How does the water mix?" Boundary processes are clearly important to establishing properties of the ocean. There are at least three boundaries: top, bottom, and sides. Air-sea interaction at the top boundary is complex and coupled to boundary layer meteorology on one side and to wave processes on the other side of the interface. Benthic boundary layer interactions at the bottom between the fluid and the sediment have consequences for each. Shore processes at the sides are looked at with interest by developers in beach communities but have consequences for physical oceanographers and geologists as well. All of these boundaries influence the physical characteristics of the ocean and present a challenge for measurement. Internal stratification and mixing are in some respects harder to observe yet accurate modeling of these is essential to rationalize the distribution of properties observed in the ocean. In fact, the classical mechanical description of fluids while helpful to understanding the physical behavior of the ocean is not sufficient and must be extended by observation.

Geophysics has relied on instrumentation at least as much as physical oceanography has to reveal seismic activity, sub seafloor stratigraphy, and most recently, hydrothermal activity at spreading centers. The revolution in marine geology starting with seafloor spreading and now incorporated in plate tectonics is based on models stimulated and underpinned by geophysical observations with oceanographic instruments (geologists might say with geophysical instruments).

Marine chemistry has recently migrated traditional shipboard chemical analyses to in situ measurements for long term, unattended observations of chemical properties of seawater. But tracer chemicals such as radionuclides and Freon have been sampled with oceanographic instruments since their discovery as tracers. While not strictly oceanographic instruments, the ion probe and the accelerator mass spectrometer are two powerful tools for dating geological samples, a key to understanding earth processes. MEMS chemical analyzers on a chip are in development and should automate the process for in situ instrumentation.

Marine biology, where there is a long tradition of naturalistic taxonomic and behavioral research, is now demanding sophisticated imaging, in situ analyses of plankton, and optical spectral measurements among other instrumentation challenges. Acoustic techniques for non invasive sampling are gaining favor for fisheries research. Medical instrumentation is being used with marine mammals for physiological studies. Perhaps most exciting are whale studies carried out with fish tags that are small instrumented probes carried by a whale during a dive and recovered later.

In the field of Ocean Engineering it is no surprise that oceanographic instrumentation is central. Vehicles, communication, imaging systems, and seafloor tools are oceanographic instrument systems. In fact, a large segment of oceanographic engineering is instrumentation.

While oceanography is still relatively new as a science, it is a growing field accommodating the more traditional disciplines of physics, mathematics, biology, chemistry, geology and meteorology as well as the engineering disciplines of electrical, mechanical, and civil engineering and computational science. Within this growing field is a sub discipline of instrumentation. Instrumentation serves all the parts of oceanography and several of the traditional disciplines as well.

I Capturing the Signal

Outline

Physical variable, measuring
Sensor, sampling
Conditioning
Data logging/transmission
Observer
Personal experience

Measurement

The universe is running in real time. That is to say there are many things happening all at once all over the place. Measurement is the art of restricting this immense variability to something manageable. Generally one restricts the spatial extent to a single small region or a line, possibly to a plane. Next one restricts the time interval to a moment in time or to a period that is brief enough to be observed in at most an experimentalist's career. Finally one restricts the variables that are observed to a few such as temperature or temperature, salinity, and pressure. It may seem that the reason to freeze this moment in time is for subsequent analysis, over and over again with different analytic tools. The increase in storage capacity in data loggers for example encourages one to sample longer, over greater space, and of more variables than one might have several years ago. But measurement or sampling is more than freezing the moment; it is reducing the variability to something that can be comprehended. If a great deal is sampled, it may postpone this reduction in complexity to a later time. This is not to say that one should not sample as much for as long over as great a volume as one can, but one should observe that there is still the need to reduce complexity.

Measurement is distinct from sampling. Sampling implies removing the signal from the environment. Measurement implies putting some kind of scale on it. If the term sampling is used to constrain the observation in the time or space domain, then there might be several measurements in a sample, perhaps averaged, or there might be a line array of sensors whose individual measurements are summed to create a sample. This distinction is significant when spectra are considered because there is no resolution at scales less than the sampling interval. One measurement per sample is the simplest kind of sampling and will suffice for now.

Coming back to the universe running in real time, there are one or more physical variables that will be measured. Consider just one variable to start. This might be a physical quantity like position of a particle or pressure at a point. It might be a more complex variable like sound or velocity. It could be something that relates directly to a measurement or that is indirectly related to a measurement. In the latter case, the measurement is a proxy for the variable. Some examples may help to illustrate this.

In particle imaging velocimetry, PIV, a photograph is taken of a sheet of light passing through fluid with particles in it. A moment later, a second exposure on the same photographic detector is made and the distance that each particle moved is measured on

the photograph (or on the video image or other medium upon which the picture was taken). The displacement of the particle in the universe running in real time is converted into a displacement of a spot on the photograph where it can be measured as a Δx and a Δy. Distance in the real world becomes distance in the photographic sample and the measurement is directly proportional to the variable.

In the case of pressure, there is no convenient way to capture pressure directly so a proxy for pressure is used, typically the displacement of a pressure-sensitive mechanical element. One such pressure sensor is a diaphragm strain gauge. In this sensor, the pressure causes a disc of elastic material, titanium metal for example, to bend and strain gauges on the side of the disc opposite to the fluid measure this displacement or bend as a change in resistance, which in turn is an electrical signal that is measured. We are very good at measuring electrical signals so this kind of a pressure sensor is a transducer, transforming a physical variable, pressure, into an easily measured electrical variable, resistance. The resistance is not the same as pressure. And the transduction of pressure into resistance is imperfect in many ways. To name a few, there are temperature coefficients, hysteresis, response time, Bernoulli port effects, Johnson noise. But after the transduction of the pressure signal into a resistance, the information about the real world is over and whatever uncorrected errors might be in the sensor are now in the sample.

Sound is an even less directly measurable variable. Typically a sensor of sound is a kind of pressure transducer called a hydrophone but its electrical output signal is manipulated further before storage since the sound is multivariable containing a time series of pressure fluctuations at a minimum. The manipulation is generally called signal processing or signal conditioning and makes the later steps more economical in power, data volume, and security. For example, the amplitude and duration of several frequencies that are present in the signal might be measured and presented as output from the sensor. In this case, the sensor is not simply a pressure sensor or hydrophone, it is a spectrum analyzer. There is no good excuse for loss of information after extraction of the real world signal by the transducer. There is certainly loss of information by the transduction itself. So for example there might be directional information in the sound signal in the real world that is lost by a single hydrophone. There will nearly always be some limits on dynamic range of the sound that a system including a spectrum analyzer will impose. Electronic noise will be introduced. But the real limit is that a spectrum as a function of time is still not sound itself. Even a very careful sampling of the pressure at a rate several times faster than any variations in the pressure associated with the sound will still not perfectly capture the original sound. It may come close as the digital music industry maintains.

Velocity is also a variable that is not even quite a physical quantity. Displacement of a fluid particle in a short interval of time is the velocity of that fluid particle. But the next fluid particle may not move the same distance in the same interval of time. So there is a fluid velocity field that is to be sampled and there needs to be some average over a suitably small volume to define the velocity at a point. Here the transduction of fluid velocity to a more convenient measurable presents many options and there are several current measurement technologies exploiting each of these that will be

explored later. However, a captured variable, velocity, is only representative of the variable, fluid velocity in the real world. It has restrictions on its range, resolution, noise, and accuracy that will not be easily related back to the velocity in the real world after sampling.

Conditioning

These examples show that transduction of a real world physical variable into a signal that can be measured is imperfect and may miss some characteristics and introduce other characteristics that are difficult or impossible to correct after sampling. But the sensor produces a signal that is no longer subject to degradation. At least there is little excuse for a signal to suffer degradation after being sampled in the sensor.

Conditioning of the signal from the sensor is the next step. This may include amplification, digitization, or band pass filtering. Although these are terms well known in electronic signal processing, they also have analogs in other disciplines. Chemical amplification may be an increase in color absorption by reaction with a reagent. Digitization could be a digital spectrophotometer reading. Band pass filtering could be mixing into a sample volume that is a certain size, not too small or too large. Biological amplification could involve concentration of cells by water removal, sieving, or filtration. Digitization might be counting of pulses representing passage of a cell. Or it could be a digital image of an organism. In these examples, conditioning is a step done after the sensor has produced a signal that is more easily processed. There should be little or no degradation of the signal, loss of information if you will, in the conditioning step. Implicit in conditioning is the storability of the result. A digital representation of the signal can be recorded or stored on a computer disk or on tape. Until the medium is no longer supported, generally a shorter interval than the lifetime of the data on the medium, it can be recovered unchanged in any way.

Conditioning is a broad category. Clearly, conditioning can include post processing of data in many ways. In the context of oceanographic instrumentation, it is useful to consider processing that takes place before recording or data transmission. Conditioning is often a major part of an oceanographic instrument. In the STD or salinity, temperature, depth recorder, primary sensors of conductivity, temperature, and pressure produced electrical signals that were conditioned to form more scientifically useful information. The arrangement of electronic amplifiers in the STD profiling instrument combined conductivity with temperature and pressure to produce a signal proportional to salinity. In the STD conditioning, signals were combined but little or no information was lost. In principle, a reverse conditioning process could have restored the signals generated by the sensors of conductivity, temperature, and pressure. In 1965, when Neil Brown developed the STD, there were no suitable microcomputers to do this conditioning so the electronic circuits he devised did the processing in analog mode. That is no longer customary except perhaps for filtering. Nearly every instrument with sensors that produce an electrical signal has a micro computer in it for conditioning. These are used for scaling and offsetting the signals based upon calibrations done before the sensor was deployed.

Averaging is a simple process that loses information but reduces the volume of data to be stored or transmitted. It is acceptable as a non degrading conditioning process if it simply removes sensor noise without reducing signal from the environment. For example, a Doppler sensor of fluid velocity has a large single measurement ambiguity due to the short burst of echoed signal from each range bin. This is not uncertainty in actual flow velocity, just in the sensed measurement. Averaging many such single measurements reduces the uncertainty of the sample and is a justified conditioning process since it doesn't degrade the information available in the measurement.

Statistical conditioning of a higher order than averaging is an acceptable conditioning process if it aids in making the sample more useful or compact. If the measurement is already dense, meaning that it does not contain sensor noise, and it is conditioned to a more compact form, information will be lost. This is conservation of information; you can repackage it but it will still be as large in a conditioned form. It may however be in a more useful form. The STD conditioning was especially useful at a time when PCs and even minicomputers were not yet available. Converting from conductivity to salinity was a laborious process with tables and calculators. A direct reading of salinity was preferable to a direct reading of conductivity although the information content was the same.

Data Storage/Data Transmission

If conditioning is not to reduce the information, storing or transmitting the data is really not permitted to destroy information. In the case of digital data on tape or magnetic or optical disc, loss of information is generally a disc crash or physical damage to a tape or platter. Fortunately, the digital archiving industry is dedicated to reducing these problems to a negligible level. For oceanographic instrumentation purposes, the adaptation of standard techniques to low temperature, arbitrary orientation, and possible high or low humidity is the main challenge. Until recently, low power was also a major challenge but laptop and notebook PCs have reduced this problem by their demand for low power storage media. Development of flash memories, low power hard discs, and static RAM storage devices in large capacities has made this problem more one of selection than design.

The problem of data transmission is not so easily put aside. Zero loss of information is the goal for data transmission as well as it is for data storage. But transmission channels are still bandwidth limited except possibly for optical fiber. They become even more limited if lossless data transmission is a requirement. This is thus not a solved problem and conditioning may play a role in data compression before transmission to remove information that is not relevant to the final observer. This data compression may permit a lossless transmission of the essential information.

Observer

The final observer is part of the oceanographic instrument system in the highest sense. Until information has been understood by an observer, the sequence of steps in extracting a signal from the universe running in real time to delivering it to the observer is not finished. This is important to understand. It is particularly important to consider the final observer when data transmission bandwidth is limited. Deep space probes or

Martian landers have a bandwidth limited data channel. In such cases, a lot of data compression is warranted. An extreme case is when a two outcome hypothesis is to be tested and a single bit of data return is sufficient. In general we have moved away from such scenarios and try to return as much true information as the data channel or recording capacity will allow. After all, the observer may change the target after the data have been collected and by reprocessing at a more primitive level, may ask a different question. This is possible only if the conditioning before storage or transmission was information conserving.

Observers are, in the end, one of the most bandwidth limited components of an instrument system. In real time data can be absorbed at about 300 baud by reading, about 1200 baud by hearing, and about 9600 baud by vision. Some special tasks like image recognition are possible at a greater baud rate but most are slow compared to the capabilities of modern instruments. Of course with post processing, selection can compress the data stream to one of these more limited observation channels and yet do it again with another algorithm and yet again with another. Considering these multiprocessing paths, it is not unreasonable to ask if data compression into every path that can be conceived of cannot be done on board the instrument and each result stored or transmitted at a much lower bandwidth than the lossless information transmission that post processing would require. There has been very little innovation in this direction to date. It is a research opportunity.

Personal Experience

Salt fingers are double diffusive convective cells that transport salt vertically without a proportional transport of heat. As a potential mixing mechanism for the world oceans where warm salty water overlies cooler fresher water (most of the temperate and tropic ocean have such structure) the efficient and ubiquitous operation of salt fingering offers a mean of establishing fine structure, a deep thermocline, and the TS characteristics of all the world oceans (curved rather than straight lines). But in 1970 their existence in the ocean was only a theory unsupported by any direct observations.

This presented a measurement problem for me, a recent recruit to oceanography from upper atmospheric physics and molecular spectroscopy. The apparent signature of salt fingers was a regular array of counter flowing fluid columns, the descending column containing higher salinity than the ascending column. Since the initial problem was one of detecting their existence rather than measuring the total transport of heat and salt, anything that revealed their unique signature was a suitable measurement. Possible measurable quantities included fluid velocity, temperature contrast, conductivity contrast (because the saltier fingers have higher electrical conductivity), optical refraction, and patterns, frequencies, or other statistical signatures of any of these. I selected the index of refraction differences due to slight mismatches between the salinities and maybe the temperatures between the descending and ascending columns. Calculations indicated that these differences were too small to detect from a single finger but that the regular array would focus collimated light if it passed along rows of fingers. In fact, this property had been utilized to study sugar fingers, a laboratory analog to salt fingers, wherein slowly diffusing sugar in an 8% solution overlying salty water of a 12% density simulated warm water with slow diffusing salt overlying cooler fresher water in the ocean. In these tank

experiments, a square array of counter flowing columns formed, exchanging salt (heat in the ocean presumably) laterally but less sugar (salt in the ocean) to generate a net buoyancy release. This source of power, the release of buoyancy by differential diffusion, overcame viscosity in the fingering region and then broke down at the edges of the several centimeter thick fingering region into turbulent mixing in the layers on either side. In this generation of mixed layers adjacent to a fingering interface, the laboratory sugar fingers suggested the newly observed layer and interface fine structure of the thermocline that was being revealed with the newly invented STD. Along the plane of rows, light in the laboratory was alternately focused and diverged to leave a shadow of the fingers on a sheet of paper or a ground glass screen on the tank. I determined to form a similar shadowgraph image in the ocean from horizontally directed collimated light that I would photograph with a movie camera from a slowly sinking vehicle that I borrowed from a colleague.

The vehicle was called Autoprobe and the optical salt finger detector was incorporated in the framework of Autoprobe along with an internally recording CTD, the latest version of the STD that could make profiles of temperature and salinity. For this chapter, the vehicle is not relevant but the OSFD, as the optical salt finger detector was abbreviated, was the sensor of the physical phenomenon of salt fingers. Fingers were known to be easily disturbed in the laboratory so the sensor for the ocean was made to be as little disturbing as possible (and thus the slowly sinking Autoprobe vehicle). This disturbance of the thing to be measured is a common worry with sensors since one would like to think that the thing that was measured represents the state of the thing without the presence of the sensor. In any case, the OSFD produced a light pattern as its measurement of the index of refraction in the optical path. More precisely, it produced a pattern of light that was related to the arrangement of optical inhomogeneities in the optical path. This is where the sensor captured the signal after which deterioration of the measurement was minor and any deterioration was not justifiable.

The images were stored on photographic film. Initially this was 16 mm movie film that ran through a custom built camera. In this system a ground glass screen upon which the salt finger images were displayed was photographed. An array of 16 light emitting diodes around the screen encoded a record number to allow the images to be correlated with the temperature and salinity profile logged digitally and identified with the same record number. Later a Super 8 mm movie camera was used with a ground glass screen and digital numerals to display the record number. In the final stage of the OSFD, a digital image capture replaced the photographic film. In this system, the image was formed on the charge coupled device of a CCD video camera and logged to an 8 millimeter video tape recorder. Synchronization with the CTD data was done by recording the CTD record number on the audio track as an RS232 signal. When playing back the 8 MM video tape, the audio output was delivered to a “glass teletype” or a computer running a terminal emulator program and displaying on a monitor. When the video tape was advanced single frame by single frame, the audio recording of record number was endlessly repeated to scroll up on the monitor until the next frame.

With the OSFD, vertical bands spaced about 1.2 cm apart and having a vertical extent of 10 to 30 cm were captured at steps in the temperature profile only 30 to 50 cm

thick. The highest contrast pictures were obtained in the Mediterranean Outflow at 1260 meters depth about 250 miles west of Gibraltar. The warm salty Mediterranean water overlies the cooler fresher North Atlantic midwater and a staircase of layers 30 to 60 meters thick there are separated by thin interfaces where the temperature changes by 0.15˚C and the salinity by a comparable 0.010 psu (practical salinity units). The fingers, or at least the vertical bands, were found on those interfaces and not between. These observations made the case for the existence of salt fingers in the ocean. With acceptance of their existence, models for the differential transport of heat and salt were also accepted and seen to explain the curved TS characteristics of all the world oceans. There remained, of course, many questions that this set of observations could not answer by itself. How common is double diffusion? Where does it occur and what competes with it? What about shear; does it destroy the square packed array? In the presence of shear can a slab structure replace the square packed array and can salt slabs replace salt fingers?

There were other sensors of salt fingers that were tried including helically descending and obliquely towed fast thermistors. The response time of even fast thermistors is marginal for detecting a finger 1.2 cm in diameter. Spectral studies of the thermistors signals have indicated salt fingers where they were expected. Micro conductivity has been proposed as a sensing modality because conductivity probes are not troubled by a thermal mass or its conductive equivalent and are therefore faster in principle. The problem is making them physically tiny enough to spatially resolve a 1.2 cm finger. Most recently, acoustics is being re-investigated with a better understanding of acoustic scattering from thermal microstructure. It turns out that optical index differences between descending and ascending fingers may disappear to first order. Had I known this at the start, perhaps by a more exhaustive search of the literature on index coefficients, I might have abandoned the shadowgraph technique before I started and missed the discovery of salt fingers in the ocean. My advice is to try things, even if they might seem doomed, because sometimes there are things that you don't know that make something work when the state of understanding at the time says it won't. My suspicion in the case of salt fingers and optical index calculations is that even when the average indices of the descending and ascending fingers match, the difference in shape of the temperature and salinity distribution across a finger allows an optical index mismatch at the edges to remain.

II Sensor as Transducer

Outline

Resistor - strain gauge, thermistor
Electric potential - thermocouple, piezoelectric transducer
Current source - polarographic electrode, photomultiplier
(Operational amplifier)
Conductivity - inductive conductivity cell, porosity meter
Photodiode – radiometer, position sensor, transmissometer
Photodiode array – camera, particle sizer
Personal experience

Resistor

Many sensors are at heart simple resistors. The change in resistance with change in the variable to be sensed is what makes them a sensor of that variable. Resistance is the ratio of voltage across the resistor to the current through the resistor. Ohm's law applies to an ideal resistor as E=IR where E is voltage (electromotive force) in volts, I is current in amps, and R is resistance in ohms. This means that the voltage across a resistor can be a measure of the resistance when a known current is made to flow through it. The digital ohmmeter uses this principle to measure resistance. A battery in the ohmmeter drives a current through a constant current source and this current is provided at the test leads of the ohmmeter to be applied to the resistor. Then the voltage between the leads is measured by the digital voltmeter that is at the core of the digital ohmmeter and this digitized voltage is displayed with appropriate scaling as resistance. The scales are changed for different resistance ranges by varying the constant current from the source and applying different scaling in the display.

Resistances are one of the most precise electrical standards we have because ratios of resistance can be made with high precision. This is done in the laboratory with a bridge circuit, either a Wheatstone bridge or a Kelvin bridge, and the bridge is balanced by adjusting a set of precision resistors until a detector of inequality between arms of the bridge can no longer detect an imbalance. When this condition is achieved, the unknown resistance is equal to or proportional to the precision resistors, independent of the voltage applied to the bridge. Precision resistors can be made from materials with a low thermal coefficient so that they can be calibrated themselves in a bridge against standard resistors, traceable to a standards laboratory and remain well known even at a temperature different from that at which they were calibrated. In oceanographic instrumentation the bridge circuit for measuring resistance is used but rarely is a set of standard resistances used to achieve balance. Rather, the imbalance is measured with a less precise voltmeter and this is related to the resistance. But the response of the sensor to the signal is in general not linear and typically the voltage measured in the resistance bridge is related to the physical variable without explicitly determining the resistance. Two examples are the strain gauge bridge and the thermistor.

Strain, the deformation of a material under stress, is typically a small displacement. When strain is a measure of another variable, like pressure in a diaphragm

pressure gauge, the material under stress is not allowed to yield or become plastic and permanently deform lest its calibration change. This limits the strain to about 2% for most metals. Thus in a length of 5 mm, the displacement due to external force will be not more than 0.1 mm. The good news is that a resistor glued to the stressed metal of the gauge can be caused to deform by 2% as well and change its resistance by 2% if stretched along its axis. This resistor is a strain gauge, made by etching a somewhat resistive metal in a pattern that will elongate when stretched in the designated direction and become thinner so its resistance increases in proportion to its strain. The strain gauge is very thin and does not affect the mechanical properties of the gauge to which it is glued to any great extent. But that is not the end of the strain gauge. There are generally multiple resistors arranged in a pattern such that two are stretched when the gauge is strained while two are not stretched or are even compressed when the gauge is strained. These four elements are combined in a bridge circuit to maximize the imbalance when the gauge is strained but to cancel the thermal effects when all four elements are heated equally. A specific example may illustrate this. The sensor is a load cell for measuring the tension in a cable.

A load cell is a strong but instrumented link that can be placed in a load carrying position in a mechanical system. For measuring tension in a cable, the load cell might have a clevis at either end that is pinned to a swaged terminal on the end of the cable above it and below it. Internally, the load cell is a bar of metal between the clevises to which a strain gauge is glued. There may be additional conditioning electronics inside the load cell or there may simply be four wires from a bridge of four strain gauge elements in the cell. The strain gauge elements themselves will be glued down with two of them aligned to sense the elongation of the load cell bar. The other two will be glued down across the bar to not increase their resistance when the bar is stretched. Figure 1 shows this strain gauge bridge schematically.

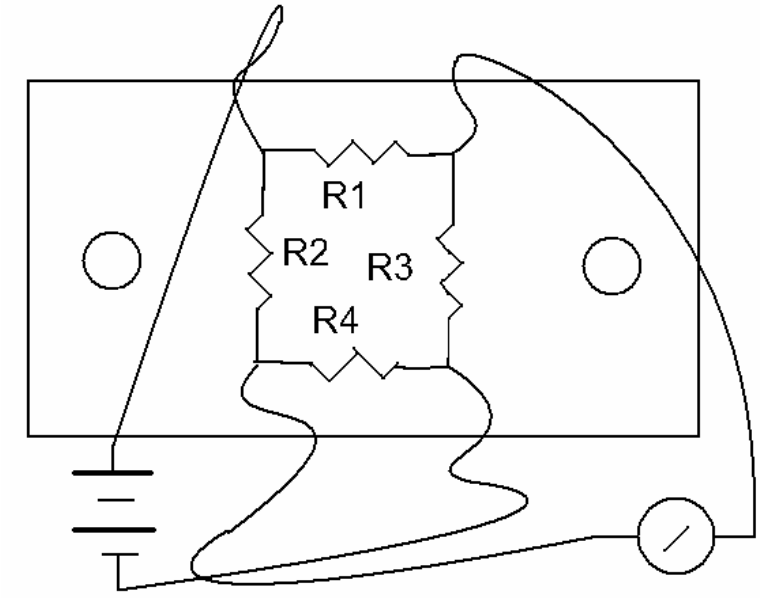

Figure 1: Strain Gauge Bridge

In the strain gauge bridge the battery supplies the voltage across the bridge and the meter reads the voltage across the other junctions. If all four resistors are equal, there is no voltage at the meter. But if the bar is stretched, R1 and R4 will increase their resistance and a positive voltage will be read at the meter. The analysis of this bridge, like other bridges, involves simple circuit analysis. Kirchoff's laws state that the voltages around a closed loop sum to zeros and that the currents into a node sum to zero. With two loops and two nodes, a set of three equations can be set up and solved. If the battery voltage is V, the current through the loop of R1 and R3 is i_1=V/(R1+R3). The current through the other loop is i_2=V/(R2+R4). No current goes through the voltmeter in this idealized circuit so the voltage at the positive terminal of the voltmeter referenced to the negative side of the battery is i_1 R3 while the voltage at the negative terminal of the voltmeter relative to the negative side of the battery is i_2R4. Across the voltmeter, the voltage is Vm= i_1 R3- i_2R4. Solving these by substituting i_1 and i_2 gives Vm=V(R3/(R1+R3)-R4/(R2+R4)). Forming a common denominator, Vm=V(R3(R2+R4) -R4(R1+R3))/((R1+R3)(R2+R4)) or Vm=V(R3R2-R1R4)/((R1+R3)(R2+R4)). Typically all four arms of a strain gauge bridge are equal to begin and the change in resistance is a small fraction of the total resistance as befits a strain of about 2% maximum, so retaining only the numerator and letting the resistances be R, Vm=V(R^2-R1R4)/$4R^2$. When the bar is stretched, both R1 and R4 increase their resistance so Vm becomes negative. But if all four resistances change as they might with a change in temperature, there is no change in voltage or in sensitivity. The exact value of the resistance is also not important except for the power and measurement circuitry. These characteristics make a resistance bridge an attractive way to handle a resistor type of sensor.

The thermistor is the second example of a sensor that is a resistor. Certain semiconductor materials are used in amorphous form to make a temperature sensitive resistor called a thermistor. In fact, most electronic resistors have some temperature coefficient but these are now mostly metal films with a slight negative temperature coefficient and can be selected for a very low coefficient, less than 60 ppm (parts per million) per degree Celsius. Until about 1950 carbon composition resistors were most common and are still in use today. Carbon is a semiconductor that has lower resistance at elevated temperature, unlike a metal, and carbon resistors can be used to measure temperature. Proper thermistors have a larger temperature coefficient and are made in a few compositions that are well characterized to permit a single point calibration or at most a small number of points to determine their resistance as a function of temperature. Thermistors are made of high resistance or low resistance material and are constructed in large, small, tiny, and nearly microscopic beads or other shapes with electrical leads and generally an insulating covering of epoxy or glass. The measurement of temperature with a thermistor offers three lessons in instrumentation. Before these lessons, the measurement of the thermistor resistance will be addressed. Recall that the physical variable is temperature but the measurable signal will be resistance and the thermistor as a sensor is the transducer to change temperature into resistance.

Most commonly, the thermistor is put into a bridge circuit like that of Fig. 1 where R4 is the thermistor and R2, R3, and R1 are fixed precision resistors. These are again typically all the same value and match the thermistor resistance at the midpoint of

the restance it has over the range of temperatures it is expected to measure. This is not the same as the resistance at the midpoint of the temperature over which it is to be used. When these resistances are equal, the reference voltage from the junction of R2 and R4 is V/2. Then the voltage measured is Vm=V (R4/(R1+R4) -1/2). Now there is some good news that won't be discussed for several more chapters, but if the voltage is measured with a digitizer, there is a good chance that the voltage driving the bridge, V, can be the reference voltage of the digitizer which removes sensitivity to this value, V, from the equation. That makes the measurement of resistance purely ratiometric with the fixed resistors. The ratio is not a linear function of thermistor resistance, R4. On the other hand, it isn't really resistance that is wanted; it is temperature. The resistance is low for high temperatures and varies only slightly with temperature. At this part of the temperature range, the voltage measured approaches Vm~V (R4/R1 -1/2) while at low temperature where the resistance changes a lot with temperature the voltage measured approximates Vm~V (1/2-R1/R4), an inverse relationship with thermistor resistance. The net effect of this is to make the voltage output less extreme than the resistance change. This permits a better allocation of dynamic range and resolution in the thermistor measurement of temperature. It is customary to fit a fourth order polynomial to a set of calibration points (at least five points but 10 or more is better) covering the intended temperature range of the sensor. A more efficient fit can be made with a logarithmic polynomial. In either case, the separation of the bridge resistance from the thermistor temperature coefficient is ignored and the temperature to resistance ratio is the only calibration required.

The first lesson about measurements that can be drawn from the thermistor is the influence of the measuring process on the thing to be measured. The thermistor is a bead of material that is heated by the passage of electric current through it. This changes its temperature so that the temperature inferred from the resistance measurement isn't the temperature of the environment. Yet without the electric current, the resistance can't be measured. This is a self heating problem and while most severe in the thermistor it is common to many other sensors. The current must be made low to minimize this effect. Power dissipation is what heats the thermistor, power being watts equal to the product of volts across the thermistor and current through the thermistor. From ohm's law, $W=I^2R$ where W is power in watts, I is current in amps, and R is resistance in ohms. This can also be written as $W=E^2/R$ where E is the voltage in volts across the resistance. In the bridge circuit of Fig. 1, the voltage applied to the bridge is partly across the thermistor and by lowering the voltage, the power dissipated in the thermistor can be lowered. This however lowers the voltage at the measuring circuit. Doing so may compromise the measurement precision if the output voltage range becomes a small fraction of the digitizer range. So another approach is to increase the resistance of the thermistor and of the other bridge resistors. This increases R and reduces the power dissipation without lowering the output voltage.

Increasing the resistance of the thermistor and the other resistors in the bridge to reduce self heating brings us to the second lesson in measurements. The voltage measuring circuit may influence the voltage in the bridge. Ideally the voltage measurement draws no current and is accomplished with no power taken from the bridge. In practice there is some influence of the measuring circuit on the bridge and this

influence is greater when the resistance of the bridge is higher. Perhaps of greater consequence is the increase in electrical noise as the resistance in the circuits, both in the bridge and in the voltage measuring circuits, is increased. This will be discussed in a later chapter. But for now, this must be countered with averaging and this reduces the ability of the thermistor to measure rapid fluctuations in temperature such as those associated with turbulent dissipation of heat. The speed of the measurement is in conflict with the accuracy of the measurement in two ways, the first is the self heating and the second is the electronic noise.

The speed of the thermistor is related to the size of the thermistor. A tiny bead has less thermal mass than a large bead. And even though the tiny bead has less surface area for thermal conduction than the large bead, the ratio of surface area to volume favors the tiny bead. Until its diameter approaches the thickness of the viscous fluid boundary layer, the frequency response of the thermistor varies inversely with the diameter. And this brings us to the third lesson in measurements. The smaller thermistor is faster but has greater sensitivity to self heating. So a tiny bead cannot be sensed with as much power dissipation as a large bead and the measurement must be averaged longer to reduce the electronic noise which compromises the frequency response. So the lesson is that there must be an optimal size where the frequency response limitation from physical sensor size is matched by the frequency response limit set by electronic noise in the measuring circuit. This may not be of much concern in a temperature controller for a calibration bath but may be a dominant design limitation in a salt finger sensing temperature probe.

Electric Potential

In the measurement of resistance, the electrical measurement is often a voltage measurement. The conversion of the resistance to a voltage is done in a bridge with an externally supplied voltage. But some sensors have a voltage output as their response to the physical variable they sense. In the case of temperature, the thermocouple is such a transducer. Dissimilar metals in electrical contact produce an emf or electromotive force (voltage) depending on the temperature of the junction. This is the Thomsen effect. It arises from the thermal excitation of electrons into the conduction band in the metal. Every junction of dissimilar metals has such a voltage. But in order to measure a voltage, wires of metal must connect the dissimilar metals of the junction to the voltmeter and the contacts of the wires of the voltmeter are themselves junctions. The only ways that the voltages in parts of the path that are not intended to measure temperature can be neglected is if the junctions are of the same metal (copper in the case of most electrical wires) or at the same temperature (in a cabinet or on a circuit board). If there is a temperature measuring junction of say iron and constantan where the temperature is to be measured, there will be at least one other junction and generally two: where the iron goes to copper and the constantan goes to copper. If both of these other junctions are at a known temperature such as in an ice bath, the emf between the copper leads will reflect the difference in temperature of the ice bath and the iron constantan junction. This is a true emf proportional to temperature. If it is measured in a circuit that introduces negligible electric current in the measurement (as in a potentiometer for example) there will be no self heating and thus a true measurement of the temperature in the environment of the thermocouple.

Unfortunately the thermoelectric effect is small. This means that the potential to be measured for temperatures in the oceanographic temperature range is in the millivolts. Thermocouples are the temperature sensor of choice for furnaces because the materials used can have high melting points so they will work at molten glass temperatures or with the right metals, at molten steel temperatures. In these cases, the emf is large enough to measure easily with a potentiometer. As a side note, a potentiometer is a device that generates a voltage and then compares the potential of the unknown to the generated voltage with a galvanometer (a sensitive current meter). The generator of voltage is adjusted until there is no current flow through the galvanometer and then the generator of voltage settings are read.

One need not have a single thermocouple however. If 10 or 100 thermocouple junctions are wired in series with one junction in each pair in the reference ice bath and other junction in the environment to be measured, the emf will be 10 or 100 times as great. The bundle of iron and constantan wires may be clumsy if the reference junction is far from the region to be measured but there are applications where a temperature difference is desired and then such a thermopile, as this series bundle of junctions is called, is just the sensor to use. In the case of a temperature measurement of salt fingers that calls on fast temperature response of a thermistor, there is the expectation that the fingers form at an interface between layers at different temperatures. A thermistor on a wing of a probe that is moving at a slight angle to the horizontal, slicing through layers and interfaces in the ocean can have this high frequency response while a thermopile between the upper and lower surfaces of the wing can measure the temperature gradient between the layers.

Another sensor that is inherently a voltage source is the piezoelectric ceramic transducer. Certain ceramics such as lead zirconate titanate, PZT, can be electrically polarized at an elevated temperature and the polarization frozen in when the temperature is reduced below the Curie temperature. Electrodes placed on the surface of a slab of such material exhibit a voltage when the material is strained. Conversely, a voltage can be placed on the electrodes and the material will change its dimensions slightly. Piezoceramic elements are used as acoustic transducers to send and to receive sound underwater. The output of the transducer is a voltage. Although the voltage is proportional to the pressure wave of the acoustic signal that is sensed, it is generally the frequency that is of greatest interest. The timing of events is among the most precise measurement we can make and frequency is easy to measure to a part in 10 million and possible to measure to a part in a trillion (10^9). So as a receiver of sound, the piezoceramic transducer is perfect as far as frequency goes. In amplitude, there are more uncertainties and generally, the sound pressure levels are related to the voltage through a calibration rather than through known properties of the ceramic. The piezoceramic elements are excellent transducers at audio frequencies. However at very low frequencies say tidal frequencies, they cannot be relied upon to accurately transform a strain into a voltage. The charge that is induced in the electrode by the ceramic being strained can bleed off. The amount of charge is small and even a very high resistance in the transducer mounting and the measurement circuitry cannot prevent this charge from conducting away over minutes or seconds. So piezoceramic transducers are used for

hydrophones and acoustic projectors (underwater loudspeakers) but not for pressure sensors.

Sensors as transducers need not be restricted to small elements like thermocouples. In oceanographic instrumentation sensors are often packaged instruments themselves and they have in many cases a voltage output. An optical transmissometer may have a voltage output of 5 volts for 100% transmission and 0 volts for 0% transmission with output voltage linearly dependent on the optical transmission. The characteristics of the output stage of this kind of sensor should be known such as the output impedance and frequency response but generally they present a simple task of digitization of their output voltage.

Current Source

When a device has very high impedance and produces an electric current as an output, it is said to be a current source. An ideal current source can be made as an electronic element but there are some sensors that are inherently current sources. The characteristic of a current source that distinguishes it from some other electrical output is that the current is the same no matter what the voltage is that might be imposed upon the output. There is a caveat that this voltage is limited to the range over which the device works, not an infinite voltage. An extreme case of a constant current device is a radio active source of alpha particles. The collector of the charged alpha particles will collect all of them until the voltage is made so high that they are turned back to the radio active source. The voltage on the collector will not effect the rate of radio active decay.

The polarographic oxygen electrode, an oxygen sensor, produces an electric current that is proportional to the rate that oxygen diffuses across a semi permeable membrane. In this sensor, a voltage is applied between a cathode and an anode. This voltage has to be selected but in a short range, the electric current is almost not dependent on the exact voltage but rather on the chemical reaction at the cathode where O_2 is reduced to OH^- by a stoichiometric number of electrons. At the polarographic voltage, about 1.83 volts, this reaction is the only serious contender for the cell. The voltage is too low to hydrolyze water into hydrogen and oxygen but high enough to reduce molecular oxygen. The reaction is $2H_2O+O_2+4e^-=>4(OH^-)$. This sensor has therefore a constant current output where that current is proportional to the rate at which O_2 diffuses across a membrane to be reduced. As a transducer, the oxygen polarographic electrode cell converts oxygen concentration in the environment, the physical variable, into an electric current, something that is easily measured. There is much to be discussed about the measurement of oxygen but that is another topic.

A simpler example of a constant current source is the photomultiplier. Within a photomultiplier, light striking the photocathode ejects an electron with a certain probability, say 20%. This electron is accelerated to the first dynode at a positive voltage with respect to the cathode of about 100 volts. At this first dynode, secondary electrons are ejected with an efficiency greater than unity, say 130%. These secondary electrons are accelerated through an additional 100 volts to a second dynode where again secondaries are ejected with an efficiency of 200%. After 20 stages of such multiplication the original photon is converted into a flood of electrons where the number

$N = 0.2*(2)^{20} = 2.09*10^5$. The last stage, the anode, collects all of these electrons even if the voltage is not exactly at ground potential (where the photocathode is at -2000 volts). So these electrons will arrive independent of the voltage within a range of about 10 or 20 volts. That is a constant current source. The task of converting this constant current into a voltage that might be digitized with ease falls to the electrometer circuit. This is one of the first non-trivial operational amplifier circuits that it is useful to understand. But a detour is needed to introduce this basic electronic building block.

(Operational Amplifier)

An amplifier increases the magnitude of a signal. Audio amplifiers may increase the amplitude by 10 or 100 and the photomultiplier as an amplifier increased the photoelectric signal by about 2 million. But an idealized operational amplifier increases the signal by a factor of infinity. It is not used as a simple amplifier however. It is used with feedback so that the behavior is defined by three rules. In Figure 2 the inputs to the OpAmp are indicated as In+ and In-. The output is at the vertex of the triangle. The first rule is that the output voltage is $(In+ - In-) \cdot \infty$ so that if there is any difference in voltage at the inputs it appears as an infinite voltage at the output with the sign of the input that is greater. The second rule is that no current flows into the input terminals; they are infinite impedance. The third rule is that when operating correctly there is no voltage difference between the inputs. If the positive input is connected to ground (a term for the common conductor to which all other voltages are referenced) and the anode of the photomultiplier is connected to the negative input terminal and finally a very large resistor is connected from the junction of the photomultiplier anode and the In- terminal to the output terminal, the OpAmp is an electrometer. The output voltage will be the negative of the anode current times the feedback resistor. This can be verified by applying the three rules. Rule one says that if the voltage of In- is positive at all with respect to ground, the output will go to negative infinity. But as it does so, the resistor will draw current away from the In- terminal making that terminal go negative until it is again exactly at ground potential where the output will stop being either positive or negative infinity. The second rule requires that any current that comes from the photomultiplier anode must pass through the feedback resistor since it cannot go into the In- terminal. This allows a calculation of the output voltage. It is negative anode current times feedback resistor. Applying the third law, the In- is at ground potential because the In+ is connected to ground and this is what allows the photomultiplier to function since the anode must be close to ground potential; to collect all the electrons. In a practical electrometer circuit using an OpAmp, the feedback resistor may be as must as 10^{10} ohms. In the example of the photomultiplier, if 100 photons/second arrive at the photocathode and the OpAmp with 10^{10} ohms feedback resistor is used, the output will be 100/second*$2*10^5$ electrons*$1.6*10^{-19}$ coulombs/electron * 10^{10} ohms * amp/coulomb/sec = 32 mV or $32*10^{-3}$ volts. This isn't a very large voltage but it can be measured so the electrometer amplifier with a photomultiplier may be able to measure a photon flux as low as 100 photons/second.

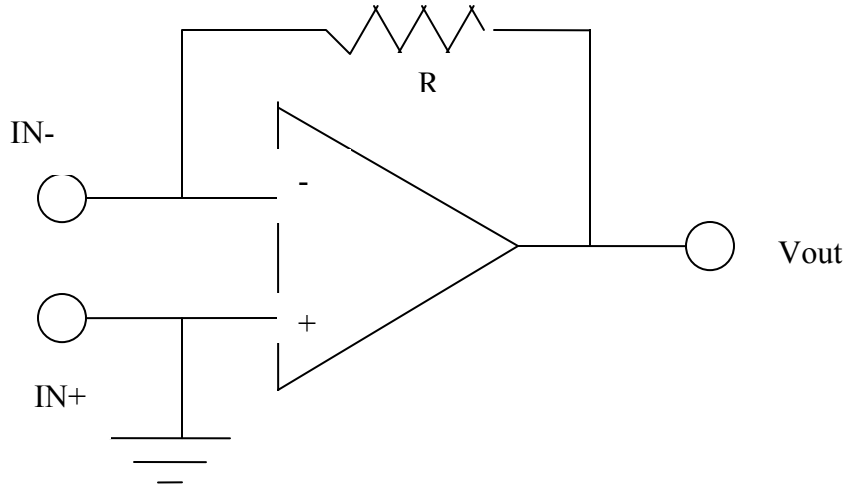

Fig. 2. Operational Amplifier as Electrometer

Conductivity

Just as a sensor may appear to be a resistor, a sensor may appear to be an inverse of a resistor and the output is conductance. This really is not very different from a resistance except in the scaling. In the CTD that measures properties of seawater, the conductivity cell is such a sensor. The four electrode cell illustrates this kind of transducer, transforming the conductivity of seawater into the simple electrical signal of conductance. An electrical current is caused to pass through a constricted cell filled with seawater from the ocean. The voltage is measured and the current is adjusted to make this voltage equal to a reference. If the current is high, the conductance is high so that the current is directly and linearly proportional to conductance. In practice the conductivity cell will polarize if the electric current flows in one direction for very long so an oscillating current is used, typically around 10 kHz. This adds complexity to the electrical measurement since phase of the applied current must be matched to the reference voltage and there may be electrical elements that introduce a phase shift that requires correction. Precise measurement of even a resistance requires care in placement of electrical contacts since these contacts have resistance and must not be included in the measurement. In the case of standard resistors, those made to calibrate other instruments, there are two leads for introducing current and two leads for measuring voltage. The contacts for the current are outside the region where the voltage is measured and there is no or very little current carried through the voltage measuring leads so their contact resistance is not included in the measurement. This so called four electrode configuration is appropriate for a conductivity cell. Two electrodes pump electrical current into the seawater in the cell while two other electrodes sense the voltage in a region of the cell that does not include the current electrodes. In this way, some polarization of the current electrodes does not effect the voltage measurement. This four electrode cell is illustrated in a configuration used in the Neil Brown Mark I CTD as Fig. 3.

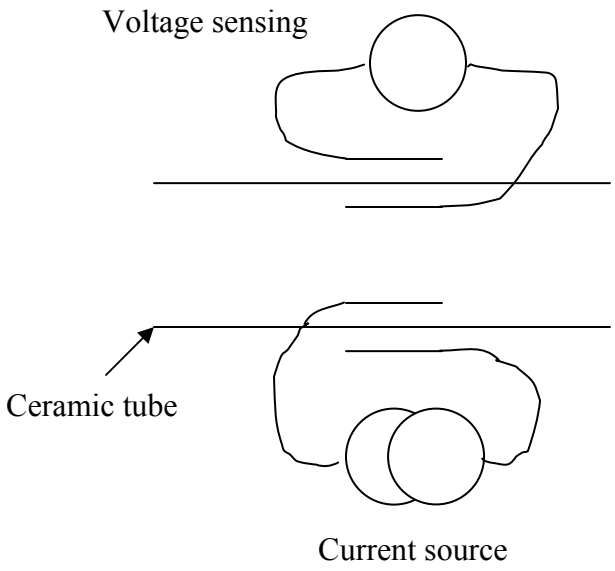

Fig. 3. Four Electrode Conductivity Cell

In the Neil Brown CTD, the adjustment of current and its reading became an integral part of the digitizer rather than having one circuit generate the current and then reading the current in a subsequent stage. The 10 kHz digitizer illustrates sophisticated instrumentation in its highest form. This will be explained in a later chapter. The sensor as transducer is here a conductivity signal, similar to what might be obtained from a standard resistor but derived from the conductivity of seawater in a fixed geometry cell.

The four electrode conductivity cell can be made small and thus measure the salinity at scales less than 5 centimeters. But the size of the cell makes it sensitive to contamination; particularly build up of lime (calcium carbonate). Periodic cleaning is necessary to retain the calibration that is inherent in the exact dimensions of the cell. In many applications, probably in most applications, the high resolution of the electrode type conductivity cell does not outweigh the burden of cleaning and recalibrating it. For this reason, an inductive conductivity cell is popular; a conductivity cell without electrodes and with a larger size that is less sensitive to lime buildup.

Recall that the measurement of conductivity in an electrode type cell requires an alternating current about 10 kHz to prevent polarization. This same alternating current can drive a transformer winding that generates an emf in any conductor that loops the magnetic field of the first winding. This is the principle of magnetic induction. Actually, the current in the primary winding is generally small because of the self inductance of the magnetic core around which the coil is wound. If the magnetic material is a toroid or donut shaped coil, an emf proportional to the voltage divided by the number of primary turns wound around the toroid will be generated in a conductor wrapped around the toroid as a kind of secondary conductor. Seawater can serve as this conductor. It passes through the hole in the toroid and around the outside of the toroid. Because the seawater has a finite conductivity, this emf will induce a current. And it is this current that will be

reflected as a current in the primary winding of the toroid. The current will flow in the seawater without benefit of electrodes and thus there will be no polarization and the troubles that such a change in electrode geometry might cause. The current that flows in the seawater also flows in the primary but reduced by the number of turns in the primary. While this current might be sensed, it is customary to generate an equal and opposite current in another winding to drive the primary current to zero and to measure that current instead. In this way, the primary winding will have essentially no current flowing through it and will not have any loss in exciting emf due to resistive losses in the wire. This is an essential treatment of the inductive conductivity cell.

In an inductive cell of outside dimension 2 cm with an opening about 0.8 cm in diameter and 1 centimeter long, the opening is modest in size to minimize sensitivity to minor fouling but it is still important to have the dimensions be stable. Since the cell may be subjected to large pressures and certainly to changes in temperature, the hole through the center is lined with a quartz glass tube sealed at one end with an O-ring. Pressure is experienced equally on the inside and outside so it is not stressed by pressure changes yet all the electrical current flow in the water must pass inside the tube. Because quartz glass has a low thermal coefficient of expansion, the cell does not change geometry with changes in temperature. Finally, quartz glass is not corroded in seawater nor is it easily scratched or worn by mild cleaning. Thus the cell constant remains constant until the liner breaks when a new liner can be installed and recalibrated or a cell constant delivered with the cell can be used for approximate values.

A second example of conductivity sensor is a porosity meter for sediment. In fact this is really a conductivity cell in which the electrodes are forced into seawater saturated sediment. In this case instead of resolving several parts per million in conductivity, as is required for salinity determinations, a resolution of several percent is required. When sediment occupies 10% of the volume or the porosity is 90%, the conductivity is 90% that of seawater. When the sediment occupies 90% of the volume and the porosity is reduced to 10%, the conductivity is reduced to 10% of the conductivity of seawater. So a conductivity sensor is a transducer for converting porosity into conductivity, a measurable signal.

Photodiode

The photodiode is a sensor that converts light into an electrical signal. More particularly it converts photon flux into electric current. As in the photomultiplier's photocathode, the capture of a photon in the photodiode generates a hole/electron pair that travels across the junction as electrical current. Not every photon creates a charge and the quantum efficiency, as the probability is called, depends on the wavelength of the photon ($e=h\nu$, $\lambda=c/\nu$, where e is the energy of the photon, h is Plank's constant, ν is the frequency of the photon, λ is the wavelength of the photon, and c is the speed of light) and the material from which the photodiode is made. There are characteristics of any sensor that distinguish the output signal from the environmental variable being sensed. In a photodiode these include the wavelength sensitivity, the angular sensitivity, and many other characteristics of the photodiode like thermal noise, dark current, and photon saturation. As a sensor, the photodiode is relatively easy to understand. But light is more complex as an environmental variable than a scalar quantity such as temperature or

conductivity. Light has color and direction as well as intensity and for each question about how much light there is a different answer may be required. For a photosynthetic cell, the critical quantity may be the photon flux in a particular wavelength band from any direction. For photographic clarity in underwater archeology it may be down welling radiation in watts per square meter over the entire visible spectrum. For global heat budgets it may be the difference in down welling radiation from upwelling radiation in watts per square meter. For AVHRR satellite temperature measurements it could be the upwelling radiation in a narrow wavelength band in the infrared in a small solid angle. Units may differ from illumination to radiance to irradiance and the integration of flux over the wavelength band accepted may be energy weighted or may be simply photon counts. The photodiode at the heart of some radiometers are photon counters. But there are radiation sensors that are weighted towards energy flux that are based upon heating of a blackened surface.

For measurements in the far infrared, photons are not sufficiently energetic to form hole-electron pairs in semiconductors. Bolometer measurements of far infrared radiation are standard for wavelengths greater than about 10,000 nm. Heating of a blackened sphere raises its temperature and increase the pressure of the gas within. If the infrared light falling on the bolometer is occulted with a mechanical chopper, the pressure signal is proportional to the energy in the infrared beam falling on the sensor.

One useful transducer is the PAR sensor for photosynthetically active radiation, a wavelength band that is absorbed by plants and used by their chlorophyll to make sugar from CO_2 and water. This measures total energy in this band by differential heating of a blackened sector versus a white sector in a disk exposed to the upper hemisphere of the sky. It is not a photodiode so it doesn't count photons. Rather it senses energy in a wavelength band and for photosynthetic efficiency estimates the energy must be converted back to photon flux.

The photodiode is much less sensitive than a photomultiplier, by the electron amplification of the dynode cascade. This is acceptable when there is a lot of light. When the light level is low but the high voltage required for the photomultiplier and its large size is not wanted, a compromise is the photo avalanche diode in which the junction in the diode is run close to its avalanche breakdown limit. This condition produces a shower of charges for each photon that creates a hole-electron pair. Although the gain in this device is not as precise as in the photodiode, it is up to 100 times as sensitive and can detect light where the photodiode cannot.

Photodiode Array

A photodiode can be and generally is made like a transistor, on a piece of silicon with etched and doped patterns and beam welded wires to the header of the device. More than one photodiode can be placed on this header. This leads to two possibilities: making a special shaped array where the photo active areas are in a particular pattern that can pick out particular features in the illumination field. The other is a raster of photo sensitive elements that can form a two dimensional image, a CCD camera chip. In the specialized array, an example is the LISST particle sizer detector array. In this array, photosensitive elements are arranged in concentric sectors of circles so that the angle of

light scattering can be distinguished and the sectors allow overlapping ranges without shorting out the elements. Larger radius elements correspond to larger scattering angles. A lead is brought out for each element of the array.

For more than a dozen or so elements, the number of leads becomes limiting and in the CCD camera array, the selection of elements is done with processing circuitry on board. CCD stands for charge coupled device. The photosensitive element transfers its charge into a field effect transistor gate, an element in the FET that is normally insulated. In a 5 mega pixel CCD array there are three photosensitive elements behind color filters to provide color information at each of 1944 rows and 2592 columns. The sequence that is used to capture an image without blurring is to empty the charge collectors by making them conductive to the base layer of the semiconductor, then after the allowed exposure time, the collected photon generated charges in each pixel are transferred to the gates of the FETs. This freezes the image and the individual photosensitive elements are the sensors that transform the light signal to electrical charges. The conditioning circuitry is on the chip because there are not enough wires nor enough chip real estate to connect the individual pixels to the circuit board. So the columns are shifted one pixel in each into a master row and then the charges in the master row are shifted out one pixel at a time as a video signal. Or there can be a digitizer on the CCDS chip so that a digital representation of each pixel is shifter out for each pixel. The moving of charge along in the columns and then in the rows is called bucket brigade charge coupling. Even after the 4500 shifts for the last element of the last row, the charge is not appreciably degraded. And the time it takes to shift all of these charges to the digitizer is the price to pay for the savings in wires. There are trade offs in the CCD array that effect the ability to freeze a color image. The three shot CCD array takes one picture in each of the three colors sequentially with less immunity to vibration since the image may shift between each shot. The single shot CCD has lower resolution but better freezing of the image.

Arrays of sensors permit patterns to be detected. This may be a different task than determining light levels. For example a pattern of bands that are aligned parallel to one another could be a signature of salt fingers in a shadowgraph detector. By using a single line array of photodiodes, a sweep of light across them may permit a laser line scan image of the bottom in murky water to be made. The principle is that a moving laser beam spotlights a single point on the bottom and that point is imaged to a single photodiode in the linear array. A moment later, the next point is imaged onto the next element of the array. When the laser has swept across the entire field of view, the linear array is read out and digitized but for each pixel that was exposed, the laser bem was restricted to a single line in the water to minimize total scattering from the water back into the lens. Yet the illumination of the point on the bottom was bright enough to get accurate albedo measurements and thus see the texture and composition of bottom material.

Personal Experience

Heat flux from the atmosphere into the sea requires turbulent transport of heat through the upper few meters of water from the very near surface. The interface is where insolation may warm the surface layer, infrared radiation into the night sky may cool the surface, evaporation from a drying wind may cool the surface, or condensation from a

warm fog may warm the surface. It requires turbulent exchange of fluid between this top few millimeters of water and the water farther from the interface to capture this temperature change. If velocity and temperature fluctuations can both be measured at a frequency and spatial scale that capture the spectrum of turbulence in this region, a direct measure of heat transport can be generated by forming the product of the fluctuations and integrating for an interval long compared to the turbulent eddy time scale. (That means you must average many individual eddy events.)

On a tower extending upward 15 meters from the seafloor to the surface south of Martha's Vineyard is a subsurface horizontal beam to which were mounted five ADV (Acoustic Doppler Velocimeter) current probes. We needed thermistors to provide the temperature fluctuation signal for this correlation. I made the temperature sensors from commercial micro bead glass encapsulated thermistors. Previous experience with glass fast thermistors that I had used bare had taught me that pressure transmitted to the thermistor also changed the resistance and the slightest touch by a finger or a piece of protective covering would break the glass. So I chose to put the thermistors in a stainless steel tube. To keep the temperature response as fast as possible I chose a fine thin walled needle for the tube and a thermistor only slightly smaller in major diameter than the inside diameter of the tube. The thermistor was 0.032" in diameter with platinum leads only 0.002" wide. The tube was 0.040" inside diameter and 0.062" outside diameter. Although the time constant for the micro bead thermistor was about 0.030 seconds (the time after plunging it into a fluid of a different temperature when the resistance reached 1/e of its final shift in resistance) the influence of the needle was expected to increase the time constant by about a factor of five. First there would be the thermal mass of the needle. While its surface area was twice that of the bead, the cross sectional area was four times as great. Second, there would be a problem of efficient conduction from the needle to the thermistor. Conduction from the needle to the bead was to be made as high as possible with thermal grease. Finally, these parts were as small as I could imagine handling.

The next problem was insulating the wires and connecting them to other wires that would be large enough to handle. Finally there was the problem of closing the tube and sealing it to a standard underwater connector. Needles were custom made and could be supplied with closed ends, welded, if required. I could not imagine pushing the 0.032" micro bead into a tube that small for a distance of several inches. So I elected to have the needle open at each end. Then I glued two lengths of enamel coated copper magnet wire together by wiping them with spray enamel and letting them dry. The wire was # 56 at a diameter of 0.014" each. I then poked the double wire through the 2" long needle. I removed the enamel at the ends of the wire by rubbing an Exacto knife along them and then applying a coating of solder to them. The platinum leads of the micro bead were cut to a staggered length to match the staggered ends of the magnet wire so that the soldered joint on one would be insulated from the other. Than a drop of nail polish was applied and allowed to dry. This was plastic when dry and could be squeezed to make the insulated part round and small enough to enter the needle. Last, the micro bead was coated in thermal grease containing zinc oxide for thermal conductivity without electrical conductivity. The wires were pulled to make the bead slip inside the needle for 3/16". It was relatively easy to solder the magnet wire to Kynar insulated #40 wire which in turn was soldered to #24 wire which was threaded through a short tube and soldered to the

#18 wire of the underwater connector. The tip end of the needle was encased in a small drop of urethane and the solder splice was covered along with the join from the cable to the needle in urethane poured into a mold. Then when the urethane had cured, each thermistor was plunged into ice water from the warm room air and the resistance displayed on an oscilloscope. The time constants were seen to be 120 ms except for two that were about 500 ms. These slow thermistors were improved by removing the urethane drop on the end and ramming in more thermal grease with a fine wire and replacing the urethane. This repaired the slow response.

Five of these 10 kilohm thermistors were tied to the side of the ADV sensors and cable connected to an equal arm thermistor bridge circuit that was driven with a 2 volt precision voltage source. The voltage across the thermistor in this equal arm bridge was 1 volt and the power dissipation was 100 microwatts. This did not overheat the thermistor although it did represent a slight warming in still air. In water however it was not a noticeable effect. The more serious potential error would have been if there was enough overheat that the thermistor became a velocity sensor, indicating a lower temperature whenever the water velocity increased. The correlation between velocity and temperature fluctuations would then be biased. This was not a problem at the 10 millidegree level.

III Noise and Limits to Measurement

Outline

Signal to noise ratio
Johnson noise
Bandwidth
Frequency response
Shannon's theorem
Personal experience
Problems

Signal to Noise Ratio

"One man's signal is another man's noise." This aphorism points out that noise is only an unwanted signal. Yet there are certain kinds of degradation of information that qualify as noise of little use to anyone. When unwanted noise is mixed with signal, the ability to recover the information is compromised. A small amount of noise is tolerable and hardly degrades the understandability of the signal. A large amount of noise can make a signal undetectable. The ratio of signal power to noise power is called the signal to noise ratio and is written as S/N and expressed in dB for decibels.

The first part of the definition is signal power. This is expressed in watts or nanowatts but in fact is rarely reduced to actual power but only kept as a ratio to a power reference. As an example, a hydrophone might produce a sinusoidal signal corresponding to a transmission of monochromatic sound 100 meters away that has an amplitude of 100 millivolts. By amplitude here, I mean the signal is V (t) =100 millivolts*sin (2πft) where f is the frequency in Hz (Hertz, cycles per second) and t is time in seconds. One then has a signal that is 200 millivolts peak to peak (V_{p-p}), an easy measurement to make on an oscilloscope. The other voltage that is used is rms for root mean square. If V (t) =A sin (ωt), $V^2=A^2\sin^2(\omega t)$, $\langle V^2\rangle=\langle A^2\sin^2(\omega t)\rangle = \langle A^2\rangle*\langle\sin^2(\omega t)\rangle = \frac{1}{2}A^2$. Then the square root of this is $V_{rms} = A/\sqrt{2}$. In the case just presented, V_{rms} =70.7 mv. Let's assume that this hydrophone with its output of 100 millivolts amplitude is driving a 50 ohm load so that the current is I (t) = 2 milliamps * sin (2πft) [I_{rms} = 1.414 ma] and the power is W = V * I = 100 millivolts * 2 milliamps * $\sin^2(2\pi ft)$ = 200 microwatts * $\sin^2(2\pi ft)$. Note that the voltage and current each average to zero but the power does not. This is part of the reason for using rms values for voltage and current since they are positive definite. The voltage averages to zero as does the current but the power averages to 100 microwatts because $\langle(\sin^2(2\pi ft))\rangle = \frac{1}{2}$, and $\langle W\rangle = V_{rms} I_{rms}$. The point here is that only the power can be usefully averaged. So instead of speaking of the voltage as an amplitude, it is customary to speak of the voltage as rms voltage and then the average can be taken as that value. This is such a useful value that digital voltmeters and even moving coil voltmeters offer on their AC (alternating current) scale a reading that is V_{rms}. We measure voltage but it is power that is conserved. One alternating voltage can be transformed into another alternating voltage with nothing more than a transformer. But power cannot be transformed into more power without an energy source and power cannot be transformed into less power without dissipation as heat or exchange for another form of energy that is stored or dissipated. When a digital

voltmeter announces that AC voltage readings are true RMS values, it means that the instantaneous voltage is squared, averaged, and the square root of the average is presented. In the simpler moving coil voltmeters, a non linear scale converts an average of the rectified voltage (generated with the non linear response of a diode) into an rms scale. If the signal is sinusoidal, the values are correct. If the signal is not sinusoidal, the approximation leads to errors.

A decibel is 1/10 of a Bell, a Bell being a factor of 10 in power. It is named for Alexander Graham Bell who certainly had reason to worry about sound power. It was probably he who noticed that an apparent doubling in loudness in sound required an increase in power by a factor of 10. The Bell is rarely used. The decibel or dB is the power ratio that is common. Now here is the stumbling block, since dB refers to power but most measurements are made of voltage, a factor of 10 in voltage is 20 dB in power. Think about it a moment; if the voltage increases a factor of 10, the power increases a factor of 100. That is 2 Bells or 20 dB. Several other dB quantities are useful to memorize. 6 dB is a factor of 4 in power. [$10^{0.6} = 3.98$] A characteristic fall off in power passed by a simple RC filter (RC refers to a resistance and capacitance in series) is 6 dB/octave. This means that when the frequency doubles, the power is quartered. If you were to try to measure this with an oscilloscope, you would be measuring voltage and when the frequency was doubled, you would see a reduction in $V_{p\text{-}p}$ to 50% of what it had been. But in power, this is 6dB. The RC filter will be looked at again in more detail later.

Returning to signal to noise ratio, if the signal is 100 μw and the power of the signal is 10 times the power of the noise, the noise power is 10 μw and the ratio S/N=1 Bell or 10 dB. Now returning to the V_{rms} and the 50 ohm load resistor, if $I_{rms} = V_{rms}/50$ ohm, then $W=V_{rms}^2/50$ ohm. That makes $Vsig_{rms} = \sqrt{(50\text{ ohm}*100\ \mu w)} = 70.7$ millivolts. How about the noise? It is 10 μw = $Vnoise_{rms}^2/50$ ohm or $Vnoise_{rms} = \sqrt{(500\ \mu w\text{ ohms})} = 22.4$ mv. So when the ratio of signal power to noise power is 10, the ratio of signal voltage to noise voltage is 3.16. In communications, a signal to noise ratio of 10 dB is good and the 22.4 mv rms noise added to the 70.7 mv rms signal is generally not an impediment to nearly perfect decoding of the information in the signal. When the signal to noise ratio is 0 dB, it is possible to get much but not all of the information without heroic efforts. When S/N = -10 dB, the situation is nearly hopeless. That it is possible to even think about recovering a signal when the noise power is 10 times the signal power is due to redundancy in the transmission and patterns that can be recovered even in the presence of massive noise. As a practical matter, the limit of communications is about S/N = 0 dB.

While still on the subject of dB, another number to keep in mind is the attenuation of signal in a simple filter at what is termed the break point. The low frequency response of a simple RC low pass filter is unity. The power out is equal to the power in. There is no attenuation. At high frequency, the attenuation is 6 dB/octave. On a log log plot, log frequency on the ordinate and log voltage on the abscissa, this is a straight line with a slope of -1. (Note that the slope of -1 is for voltage not power. A log log plot of power vs. frequency would have a slope of -2 because the power drops by 6 dB/octave or a factor of four each doubling in frequency.) If the straight lines are continued, they cross

at the break point. In an RC filter, this corresponds to $\omega = 1/RC$ or $f_{bp} = 1/(2\pi RC)$. The actual attenuation at the break point is 3 dB. This is a factor of 2 in power [$10^{0.3} = 2.00$]. The voltage is reduced to 70.7% of the low frequency voltage.

Johnson Noise

While on the signal to noise subject, the noise must be considered. External sources of noise may creep into most electronic systems but there is an inherent noise of quantum mechanical origin caused by the random fluctuations of electrons in a conductor that is called Johnson noise or thermal noise. This is the limit for noise reduction below which one cannot go. The thermal excitation of electrons is broadband meaning that it has energy at high frequencies. In fact, the noise power in a resistor is proportional to the bandwidth and the absolute temperature. The power is not dependent on the value of the resistance but it is a property of the resistor. Zero resistance, no noise power. But in any finite resistance, the noise power follows the relation:

$W_{noise} = 4KTB$

where, K = Boltzmann's constant - $1.38 * 10^{-23}$ J/K
J = Joules (watt-second)
K = Kelvin (absolute Celsius scale – degrees)
T = temperature (K)
B = bandwidth (Hz)

Introducing the specific value of the resistor, R in ohms,

$i_n^2 = 4KTB/R$

$e_n^2 = 4KTBR$

where i_n^2 = mean squared noise current (amps)
e_n^2 = mean squared noise voltage (volts).

Amplification is commonly the next conditioning step done in an electronic instrument and the process of amplification involves resistances in every practical case. This introduces noise from thermal agitation of the electrons in the resistors so that the process of amplification introduces noise. If the signal is large compared to this introduced noise, high S/N, there is little degradation of the signal by amplification. If the S/N is not large, something else must be done in this conditioning step and there are but two terms that can be adjusted: T and B. Temperature is lowered in photo detectors as a routine way to improve the signal to noise ratio. But in nearly every other case, the bandwidth is lowered.

Bandwidth

The breakpoint of a simple RC filter can be considered the bandwidth limit of a circuit and define the frequency range over which Johnson noise exists to degrade the signal to noise ratio. Frequencies below the breakpoint are not attenuated while frequencies above the breakpoint are attenuated at 6dB/octave. Integrating just a little bit into the high frequency part of the spectrum shows that the contribution is negligible. It is effectively made up for by the actual attenuation of 3 dB at the breakpoint. But bandwidth is a range in frequencies, not just the upper frequency limit. A bandpass filter is a filter that permits a range of frequencies to pass with modest or no attenuation but that attenuates low frequencies and high frequencies. If the slope of attenuation is 6 dB/octave or greater, it is appropriate to let the pass band or bandwidth be the upper

frequency breakpoint minus the lower frequency breakpoint. For radio frequency circuits the bandwidth is nearly always a pass band of this type.

While this simple low pass filter or bandpass filter is adequate for a discussion of noise, there are more complex filters that attenuate more rapidly than 6 dB/octave or have desirable characteristics for certain applications such as no phase shift, reduced sidebands, or very flat response over the pass band.

Frequency Response

Closely related to bandwidth is frequency response. A sensor with a high bandwidth has a high frequency response. This is a specific reference to a bandwidth that extends from DC or from zero frequency. Practical examples of frequency response include acoustic systems such as a telephone or a speaker system. Frequency response of a telephone is said to be 5 kHz because the breakpoint if viewed as a low pass filter is at 5 kHz. A speaker system with a cut off of 15 kHz is said to have a frequency response of 15 kHz when above that frequency the power is strongly attenuated, 6 dB/octave for a simple filter.

This simple RC filter that we have referred to can be instructive if we analyze it using Kirckoff's laws. In Fig. 4 the instantaneous current through the capacitor is $i_C(t) = (E(t)-v_C(t))/R$ where E is the applied emf or voltage and i_C and v_C are current intoand voltage across the capacitor respectively.

After the physical signal is transformed to some other form by the sensor, it can be conditioned by amplifying, filtering, correlating, or sampling. Amplifying can generally be done with little degradation of signal to noise ratio. The term signal to noise ratio defines the ratio of useful information to unwanted information. Extra information added by the amplifier is unwanted. So if it adds information, this is noise. Similarly if in addition to amplifying the signal, it loses a part of it, this reduces the signal to noise ratio. Broadband electrical noise is generated by thermal excitation of electrons in a resistor. Because electronic amplifiers have resistive components, this thermal noise is introduced by amplification. It becomes the limit of the system in those cases where the sensor signal has very low level (as in the case of radio or acoustic receivers) and care must be taken to deal with this noise source and not permit it to become worse than the theoretical limit.

The thermal noise of a resistor is called Johnson noise. Considered as a current source in series with the resistor, or as a voltage source in parallel with the resistor, two expressions for the noise can be written:

$$i_n^2 = 4KTB/R; \quad e_n^2 = 4KTBR$$

where i_n^2 = mean squared noise current (amps)
e_n^2 = mean squared noise voltage (volts)
K = Boltzmann's constant - $1.38 * 10^{-23}$ J/K
T = temperature (K)
B = bandwidth (Hz)

R = resistance (ohms)

From this it can be seen that cooling the amplifier may be required for achieving the best noise figure and this is done for certain space receivers. However for the rest of us, the only reasonable thing to do is lower the bandwidth.

Bandwidth is the range in frequency passed by the system. Analog filtering and sampling determine the bandwidth. In a communication system, the bandwidth determines how fast information can be transferred and there is a tradeoff between a fast system and a reliable one. The simplest way to limit bandwidth is with a low pass filter since the bandwidth below a certain frequency is limited to that frequency while the bandwidth above is infinite. In high frequency systems, a narrow band filter is used.

Problem Set - Noise and Sampling

1) What is the voltage noise of a 100K resistor at room temperature and a bandwidth of 10 MHz. Watch units.

2) An FM radio receiver may see an antenna with a 300 ohm impedance. (This is reactive, not resistive, but a maximum power match might use a transformer to reflect the input transistor impedance to that value. Think of a 300 Ωnoise source at the input.) A commercial FM station is ~ 100 MHz and the radio channel (not TV) is 30 kHZ wide. What signal strength is needed to get 0 db S/N? [db = $20*\log_{10}$(voltage ratio)]

3) Extra credit. A deep space probe at a range of 10 astronomical units drives its 2 meter diameter 900 MHz antenna with 10 watts. What is the signal received at a 10 meter diameter dish antenna on earth? What is a reasonable baud rate at room temperature? At liquid helium temperature?

IV Sampling and Time Series Analysis

Outline

Digital sampling
Finite time series
Filtering and averaging
Frequency domain
Spectral analysis
Windowing
Personal experience
Problems

Digital Sampling

Capturing the state of a variable is sampling and for practical purposes with electronic signals, sampling is digitization. In cases where the sample is physical as for example in capturing a marine organism, measuring and entering the length of that organism, caught at a certain depth and at a certain time of day, into a data base becomes the digitization step. Conversion to a number is the digital part of the sampling. Since the statistics of time series analysis can be applied to digital samples, even such biological observations are appropriate for the discussion that follows.

The sampling step is both a restriction of the observation of the universe and a saving of the state of at least one variable in the universe. It is a restriction in the sense that the variable state between one sample and the next is unknown; and it is a saving of the state in the sense that after digitization, the observation will not change. At least it will remain preserved barring fire, flood, or disk crash; and there is little excuse for not having back-up of digitized data. Addressing the restriction of sampling, the best strategy is to sample at twice the highest frequency of variation or more to avoid missing variations or perhaps more important, avoiding aliasing variations at a frequency not sampled into apparent variations at a lower frequency. If the variable sampled does not have a natural frequency limit, there can be an averaging process applied before sampling that is linear and unbiased to reduce the frequency of variation to one that can be sampled at twice the highest frequency or greater. Returning to the length of the marine organism, collect it in a net of sufficient size from a tow of sufficient duration that the patchiness of the distribution of the organisms is averaged before the net is closed and the population that will be digitized later is sampled. Variations of organism abundance and length at shorter intervals and over shorter distances than the net included in the sample will be lost but their variations will not bias the digitations of the sample that are made.

Sampling is often done as a function of time. Voltage as a function of time is a common electrical signal and periodic digitization of that signal becomes a time series. Spatial series sampling is equally valid as for example the scanning of a photograph or other image in two dimensions where the sampling is two dimensional and generally uniform in both x and y over the surface of the image. Sampling is not restricted to uniform intervals although analysis of series that are not uniform is a more advanced technique. In many observations, time series are obtained where spatial sampling would have been preferred but was impractical. Reynolds stress serves as an example of this.

Reynolds stress is a measure of momentum transport across stream lines in a turbulent flow and is a spatial average over a sheet in the flow containing the streamlines. However, one commonly has a time series of velocity fluctuations at a single point over an extended period. One hopefully invokes the ergodic hypothesis that if the flow is steady and homogeneous, a time series can replace a spatial series. That is what is used for Reynolds stress observations.

Finite Time Series

The digital sample is a series, for now consider a time series although a spatial series will work as well. The series is finite. It is a piece of data. A continuous record with no end is not under consideration here and where there is such a continuous series, it must be cut into pieces for the analyses that follow. Because the series is finite, there are artifacts from the ends that have effects. Two approaches are viable for limiting the artifacts. The first is to fold the piece so that the first sample follows the last sample and becomes a repeating series. Then it is again an infinite series without end and does not have the artifact. Of course it is a repeating series and does not have more information than the original piece but it avoids the end effect. The second approach is to taper the influence of the ends and this approach is called windowing and will be discussed at the end of this chapter.

A finite series has a set of points, N, and cannot have spectral content that is higher frequency than half the number of points divided by the duration of the time series. This is a consequence of the sampling theorem that two samples per cycle of the highest frequency present are required to characterize the signal. The quality of the information at the highest frequency is limited because it is really only a single example of the variable sampled at such a frequency and may not be a stationary statistical result. To increase the assurance that the variations at the highest frequency that can be extracted from the time series is stable and representative, many examples are needed that can be analyzed and the values of the variations at the highest frequency from each example must be averaged and the differences between examples expressed as standard deviation of another measure of variation.

We have a times series that is finite, a piece of data, and now we need many more before we can state with authority what the variations at the highest frequency are. But we have a solution for the highest frequency part of the variations of the process that has been sampled. We can break the finite time series into several shorter piece, each of which has the same high frequency represented but is unique in its realization. This gives us many examples that can be averaged and compared to gain confidence of the stability of the estimate. The statistical test has validity based upon the number of degrees of freedom, that is the number of independent observations (examples) in its formation. If a finite time series is divided into M pieces and phase as well as amplitude of the resultant frequency spectrum is considered, there are 2M degrees of freedom in the average of the spectra from the M examples.

From a single finite time series there can be multiple pieces that increase the degrees of freedom for the highest frequencies but each of the shorter pieces has a low frequency cut off that is not as low as it was before it was chopped up. Recall that there must be at least a full cycle of the lowest frequency that can be extracted in the piece. So if the original finite time series is broken into M pieces, the lowest frequency that can be

determined is now M times as great as it was. The consequence of this is that you may need to take a much longer time series than you thought you wanted in order to still have enough length after the original time series is broken into M pieces.

Filtering and Averaging

Filtering reduces the variability and is needed if there is no intent to obtain information about the higher frequency variability that the filtering removes. As stated several times, the loss of information by under sampling is less severe than the aliasing that occurs when the signal contains higher frequency information than is sampled. Consequently, the high frequency limit is set by the filtering before sampling and cannot be recovered by more rapid sampling or by longer sampling and increased data quantities. However the quality of the estimates can be increased by longer sampling since the finite time series can be cut apart and then the spectral estimates averaged, increasing the degrees of freedom. Increasing the sampling frequency after filtering does not provide independent information and only serves to reduce the noise of the measurement (which may be a worth while goal).

A finite time series has two easily recognized features. The mean of the series is an average that can be removed for spectral analysis. In electrical terms, this is a D.C. offset, referring to the direct current as opposed to the A.C. or alternating current part of an electrical signal. The trend is also an easy feature to recognize and remove. This can be determined by a least squares fit to a straight line. Both of these features should be removed before further analysis for frequency components since neither has any frequency content but each interferes with least squares fits to spectral components that are matched to the data. What remains after the mean and linear trend have been removed is ready for spectral analysis.

Frequency Domain

A time series without the mean and trend can be represented equally well as a set of frequency components equal in number to the number of data points in the time series. Thus there is a complementary representation in the frequency domain of a time series. Advantages to working in the frequency domain are that events are not tied to a particular moment in time and are thus generalizeable and that averaging of several realizations is possible in the frequency domain to improve the statistics. Averaging in the time domain between different realizations makes no sense except to improve the mean of the process.

Since it takes the same number of coefficients in the frequency domain as in the time domain to represent a process, there may seem to be no possibility of data compression by transposing to the frequency domain. However the interest is not necessarily uniform across the spectrum so there is a possibility of band averaging to reduce the number of coefficients required. And this is also a natural way to combine examples from different time series.

The range of frequencies representing the time series is from one cycle for the duration of the time series (reciprocal duration), then two cycles for the duration of the series, then three cycles and so forth to as many cycles as samples or one cycle per sample interval. If for example a sample is taken every 1 s for 17 minutes (1024

samples) the highest frequency in the frequency domain is 1 Hz and the lowest is 1/1024 s (~1 mHz). The formulation for this frequency decomposition is:

$$X(k) = \sum_{j=1}^{N} x(j)\omega_N^{(j-1)(k-1)}$$

$$x(j) = (1/N)\sum_{k=1}^{N} X(k)\omega_N^{-(j-1)(k-1)}$$

where

$$\omega_N = e^{(-2\pi i)/N}$$

is an N_{th} root of unity.