

Eindhoven University of Technology

Monitoring of Respiration and Circulation

Reader

Woerlee, Pierre,
Faculty of Electrical Engineering, SPS group
11-1-2022

Table of Contents

1	PREFACE	9
2	Introduction	10
2.1	Function of Monitor and Therapeutic devices	11
2.2	Differences between Engineering and Medical fields	13
2.3	Medical Decision Making and Diagnosis Process	14
2.4	Modules of the course	16
2.4.1	Module 2: Physics and Physiology Topics in the Course	16
2.4.2	Module 3: Sensors Transducers and Clinical Measurements	17
2.4.3	Module 4: Measurements of the cardiovascular and respiratory systems	17
2.4.4	Module 5: Patient Monitors and Therapeutic devices	18
2.5	Questions	19
3	Elasticity and Viscoelasticity	20
3.1	Force and Pressure.....	20
3.2	Elasticity and Material Deformation	22
3.2.1	Elastic Deformation of materials.....	23
3.2.2	The Laplace Law in Elasticity	25
3.2.3	Volume and compliance of a thin-walled elastic tube	27
3.3	Biologic tissues and viscoelastic deformation of materials	27
3.4	References	31
3.5	Questions	31
4	FLUID TRANSPORT.....	32
4.1	FLUIDS.....	32
4.2	FLUID PROPERTIES	33
4.2.1	Pathlines	34
4.2.2	Viscosity Shear Stress and Shear Rate	36
4.2.3	Numeric values of viscosity and density	37
4.3	Volumetric and Mass Flow Rates	37
4.4	FLUID DYNAMICS	38
4.4.1	Viscid flow and Flow Resistance	39
4.5	Inertance – Inviscid Flow.....	41
4.6	Flow in an elastic tube – Compliance and Tube Collapse	43
4.6.1	Viscous flow through an elastic collapsible tube	45
4.7	Inviscid flow in long elastic tubes – transmission line	47
4.8	Bernoulli's Equation	50
4.9	Turbulence	53

4.10	Tube network models.....	56
4.10.1	Analogy between Fluid and Electrical Systems	56
4.10.2	Selection of model elements	57
4.11	References.....	60
4.12	Questions	61
5	Sensors, Transducers and Measurement Modules.....	62
5.1	Sensors and Transducers.....	62
5.2	Clinical Measurement	65
5.3	Pressure Measurement.....	66
5.3.1	Mechanical Transducers	66
5.3.2	Electronic Pressure Transducers	67
5.4	Flow Measurement	68
5.4.1	Pneumotachograph	68
5.4.2	Venturi Flow Rate Transducer.....	69
5.4.3	Pitot Tube Volumetric Flow Sensor.....	70
5.4.4	Rotameter.....	70
5.4.5	Hot wire anemometer	71
5.4.6	Thermal Mass Flow Transducer	72
5.5	Blood Flow Measurement.....	72
5.5.1	Ultrasound Transducers.....	73
5.5.2	Transducer Elements	74
5.5.3	Continuous Wave Doppler	75
5.5.4	Pulsed Ultra Sound ([Blom]).....	76
5.6	Volume Changes - Plethysmography.....	78
5.6.1	Cuff based plethysmography	78
5.6.2	Whole body plethysmograph.....	79
5.6.3	Photo Plethysmography.....	79
5.6.4	PPG Transducer and Signals.....	80
5.6.5	Electric impedance plethysmography (bio-impedance).....	82
5.7	Electrocardiogram	84
5.7.1	Input protection.....	86
5.7.2	ECG Amplifiers	87
5.7.3	Additional ECG circuit Blocks	89
5.8	Temperature	89
5.9	Motion and Acceleration.....	93
5.10	Capnography	94

5.11	Pulse Oximetry	96
5.12	References.....	102
5.13	Questions	103
6	Measurements of the Respiratory System	105
6.1	Introduction	105
6.2	Measurements	105
6.3	Lung Volumes.....	105
6.3.1	Lung volume changes - Spirometer.....	106
6.4	Measuring absolute lung volumes	108
6.4.1	Nitrogen Wash-Out method	109
6.4.2	The Helium dilution method	110
6.4.3	Whole Body Plethysmography.....	110
6.5	Measuring dead space and unequal ventilation	111
6.5.1	Single breath Nitrogen washout method.....	112
6.6	Lung Compliance	113
6.7	Pulmonary Function Tests	117
6.7.1	Airway resistance.....	117
6.7.2	Forced Expiration Test	119
6.7.3	Test of lung pathologies.....	123
6.8	Diffusion Capacity	124
6.9	Blood gases and pH (Optional).....	125
6.9.1	Capnography and Pulse Oximetry.....	126
6.9.2	Respiration Rate	126
6.10	Measurements During Sleep Disorders	128
6.11	References.....	130
6.12	Questions	131
7	Circulatory Measurements	132
7.1	Measurements on the cardio-vascular system	132
7.2	Blood Pressure	133
7.2.1	Arterial Pressure-Volume Relation	133
7.2.2	Body locations for blood pressure measurement and procedures	134
7.3	Blood Pressure Measurement.....	135
7.4	Invasive Blood Pressure	136
7.4.1	Catheters with MEMS transducers	140
7.4.2	Summary Invasive Blood Pressure	141
7.5	Non-Invasive Blood Pressure.....	141

7.5.1	Riva-Rocci Method.....	142
7.5.2	Korotkoff Auscultatory Method	143
7.5.3	Oscillometric Non-Invasive Blood Pressure	145
7.5.4	Accuracy and Verification	147
7.6	Continuous Non-Invasive Blood Pressure	149
7.6.1	Applanation Tonometry.....	149
7.6.2	Vascular Unloading Technique (Volume Clamp)	151
7.7	Continuous Non-Invasive Blood Pressure – Surrogates	153
7.8	Blood Flow and Cardiac Output	155
7.8.1	Cardiac Output - Direct Flow.....	155
7.8.2	Continuous ultrasound	156
7.8.3	Pulsed ultrasound TEE Ultrasound.....	157
7.9	Indirect Cardiac Output Measurements.....	158
7.9.1	Fick Principle	158
7.9.2	Indicator Dilution Method	160
7.9.3	Model Based Methods – Pulse Contour Methods	164
7.10	Velocity Encoded Phase Contrast MRI.....	168
7.11	References.....	170
7.12	Questions	171
8	Patient Monitor	172
8.1	Monitoring Requirements	172
8.1.1	Patient	173
8.1.2	Intervention	173
8.1.3	Location	173
8.1.4	Standards of Monitoring (ASA)	173
8.2	Vital signs	174
8.3	Patient Monitor Device	174
8.4	Patient Interface, Cables and Connectors	177
8.5	Clinical Measurements	177
8.5.1	Data Processing and Parameter Extraction.....	179
8.5.2	Data Types	181
8.6	Monitor Display.....	182
8.7	Alarming.....	184
8.8	Family of Patient Monitors.....	188
8.9	Central Monitoring Stations	192
8.10	Hospital Networking and Connectivity Solutions	192

8.11	Summary	195
8.12	References.....	197
8.13	Questions	198
9	Therapeutic Devices	199
9.1	Mechanical Ventilation	199
9.1.1	Airflow to the Lung	200
9.2	Ventilator Modes of Ventilation.....	202
9.2.1	Pressure Controlled Ventilation.....	203
9.2.2	Flow/Volume Controlled Mechanical ventilation	204
9.2.3	Positive End Expiratory pressure (PEEP)	206
9.2.4	Ventilator Induced Lung Injury (VILI)	207
9.2.5	Pro-Con Analysis of PC and VC controlled Modes of Ventilation	210
9.3	Ventilator Systems	211
9.3.1	Open Ventilator System.....	211
9.3.2	Circle System.....	212
9.3.3	Gas Inlet System	213
9.3.4	Breathing Set	214
9.4	Modern PPV Ventilator	215
9.4.1	Principal Gas flow in the G5 Ventilator	216
9.5	Ventilator Modes in a Modern PPV ventilator	218
9.5.1	Synchronization of Patient effort and Support Ventilation.....	218
9.6	Mechanical ventilation process.....	220
9.7	Physiological Closed Loop Ventilation (Smart or Intelligent Ventilation)	221
9.8	References	225
9.9	Questions Therapeutic Devices	226
10	In-Hospital Patient Monitoring.....	228
10.1	Time Critical Care and High-Acuity Monitoring in the ICU.....	228
10.1.1	Shock	229
10.1.2	Therapy during shock states	231
10.1.3	Monitoring and Measurements during Shock States in the ICU	233
10.1.4	Fluid Therapy	234
10.1.5	Severity of disease Scoring systems in the ICU	237
10.1.6	Acute Respiratory Distress Syndrome.....	239
10.2	Monitoring in the ward	241
10.2.1	Rapid Response Teams or medical Emergency Teams.....	241
10.2.2	Criteria for RRT/MET calls – Vital Signs and other parameters	242

10.2.3	Early Warning Scoring System	244
10.2.4	Monitoring Technology in the Ward	245
10.2.5	Status of Rapid Response Systems.....	251
10.3	Home Monitoring	252
10.3.1	Telemonitoring - Wearable Consumer Devices and Apps	253
10.3.2	Professional wearable devices for hospital and home use	254
10.4	Questions	257
11	Patient Monitoring Technologies - Future Developments	258
11.1	Trends in Healthcare Systems	258
11.1.1	Ageing Population.....	258
11.1.2	Proportion of Chronic Diseases.....	259
11.1.3	Shortage of Staff	260
11.1.4	Growth in Medical Knowledge.....	260
11.1.5	Cost of Healthcare	261
11.2	Challenges and Issues with STATE-OF-THE-ART Monitoring Systems.....	262
11.2.1	Transitions in monitoring during hospital care	263
11.2.2	Gaps in Monitoring	264
11.2.3	Patient Worn Devices	264
11.2.4	Warning and Alarming	264
11.2.5	Connectivity.....	265
11.2.6	Cable Clutter	266
11.2.7	Utilization of Data	266
11.3	Technology Trends in Healthcare	267
11.4	Clinical Measurements	268
11.4.1	ECG and Bioimpedance Measurements.....	268
11.4.2	Pulse Oximetry.....	269
11.4.3	MEMS accelerometers and barometric pressure sensors for respiration rate, heart rate and fall detection	270
11.4.4	Blood Pressure.....	271
11.4.5	Cardiac Output.....	272
11.5	Next generation Patient Monitor Devices	273
12	Appendices for Chapter 3.....	275
12.1	Volume-Pressure relation and Compliance for an elastic cylindrical tube	275
12.2	First and second order models of viscoelastic materials (optional).....	278
12.3	Damped spring-damper system with Inertia and Mass	279
12.3.1	Dynamic models of viscoelastic inertial systems	280

12.3.2	System with a step function driving force.....	281
12.3.3	System with harmonic driving force	282
13	Appendices for Chapter 4.....	284
13.1	Navier Stokes Equation	284
13.2	Basic Network solution methods.....	285
13.2.1	Network Relations	285
13.2.2	Kirchhoff relations	285
13.2.3	Recap Basic Circuits	286
13.2.4	Parallel RC segment	289
13.2.5	Series RLC segment.....	290
13.3	Transmission Line Models	293
13.3.1	Transmission line with losses.....	300
13.3.2	Summary TLM appendix	303
14	Appendix Chapter 5.....	304
14.1	The electromagnetic flowmeter	304
14.2	Time Transit Flow Meter	304
15	Appendix Chapter 9 – Therapeutic Devices.....	306
15.1	Impact of Positive Pressure Mechanical Ventilation on Hemodynamic Parameters	306
15.2	Adaptive Support Ventilation (ASV)	307
15.3	Intelligent ASV	311
15.4	Neurally Adjusted Ventilatory Assist (NAVA).....	311
15.5	Infusion Systems [8]	313
15.6	Defibrillators and Cardio Pulmonary Resuscitation	316
15.6.1	Types of cardiac arrest.....	316
15.6.2	Defibrillator.....	317
15.7	Cardio Pulmonary Resuscitation (CPR)	319
15.7.1	Guidelines for Cardiac Arrest Treatment	321
15.7.2	Statistics of sudden cardiac arrest	322
15.7.3	Monitoring during CPR – A case study in the laboratory	323
15.8	Image Guided Therapy	330
15.9	Pacemakers [8]	331
15.10	Heart Lung Machines [8]	332
15.11	References.....	334
16	ANSWERS TO QUESTIONS	336
16.1	Answers Chapter 3	336
16.2	Answers for questions in chapter 4	337

16.3	Answers to questions of Chapter 5.....	340
16.4	Answers to Questions of Chapter 6	345
16.5	Answers to Questions of Chapter 7	347
16.6	Answers to Questions Chapter 8	349
16.7	Answers to Questions of Chapter 9	352

1 PREFACE

This course is focused on patient monitoring of the respiratory and cardio-vascular systems in a professional clinical environment. Data of the respiratory and cardio vascular systems is of prime importance to the clinicians both for diagnosis of diseases and for therapy to ensure short term and long term stability of the patient. Note that imaging and laboratory analysis of blood, urine and tissue samples are also of high importance but these modalities fall outside the scope of this course.

A patient monitor is a complex device and (basic) knowledge of many disciplines is required to master the course. Therefore the course is split in several modules that are treated in a logical order. The modules of the mandatory course material that is required for the written examination is described in the first part of the reader. Depending on the prior knowledge of students some modules may be skipped or can be briefly reviewed. Background material that is not part of the exam but that is needed for students without prior knowledge or for students who are interested in more in-depth material is presented in the appendices. The written exam counts for 80% of the final score.

The course is accompanied by a practical part which is used to give the student hand-on information on topics relevant for the course. Simple lumped element models are used to model systems and human body systems and therapy devices to give information on measured signals, its origin and magnitude. The analogy of these models with models in the electrical domain is exploited and the Multisim circuit simulation tool is used. Furthermore there is an assignment on measuring blood pressure, signal analysis of an oscillometric blood pressure measurement and comparison of the main techniques. Finally there is an assignment on an analysis of time series data of high quality monitor data of an animal test where data from an animal in a shock state must be analyzed using the Guyton model of the circulation.

2 Introduction

A patient monitor is a device or medical test for observing a biological condition, medical parameter or function of a patient over time. Historically patients were monitored by manual palpation and listening to sounds of the patient (lung sounds, heart sounds). The first monitoring methods were developed at the end of the nineteenth century, but more widespread use of electronic monitoring devices started after World War 2. Patient monitors are typically used during critical care conditions and during the disease diagnosis process, the aim is to guide and optimize therapy. During critical care there is a focus on the cardio-vascular and respiratory systems as these are most important to stabilize the patient condition and for subsequent treatment. In short monitoring provides timely and essential information on the health status of a patient to the clinician. An important function of a monitor is to warn the clinician for imminent danger in the health status of the patient (alarming function). In the terminology of industrial process analysis, monitors are sensors that provide healthcare professionals with patient-related data and information that supports those professionals in making appropriate diagnosis and therapeutic decisions.

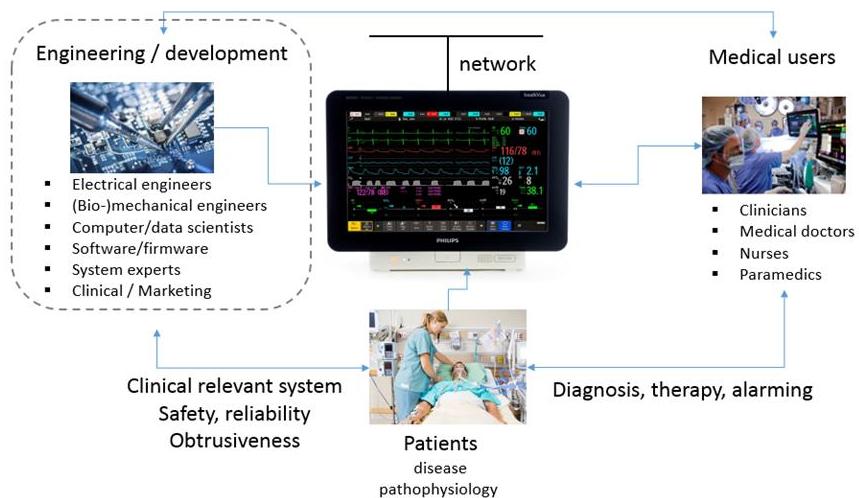


Figure 2-1 Patient monitoring environments.

A typical patient monitor and its environment in a hospital are shown schematically in Figure 2-1. First let us focus on the clinical and patient part. The patient in bed is connected to the monitor via sensors and cables, signals are processed and finally features, and parameters are extracted from the waveforms. Both the waveforms and numeric data are shown on the display. The monitor is connected to the hospital network and data is stored in the hospital IT infrastructure. During patient care clinicians use the information on the display to assess the health status of the patient and if needed adjust the therapy to stabilize the condition of the patient. The interpretation of the monitoring data and the subsequent diagnosis and therapy is described in more detail in lectures of the course. The demands on real time, quality and reliability of the monitor device are very high. Furthermore, the device should be safe for the patient and user, should be compatible with the existing workflow of the clinicians and be designed for maximum patient comfort.

The culture and way of working in the clinical environment differ appreciably from that in the engineering domain. Note that measurements in the clinical domain deviate from those in the physical and engineering sciences in many aspects (noisy signals, large dynamic range, more errors, lack of context, motion artifacts and large differences in signals in a patient in time and between patients). Furthermore, the safety demands are extreme and compliance with complex regulatory rules poses a high constraint on device development and testing.

When new products are designed it is important that engineers are aware of the clinical user needs, clinical and other requirements and clinical workflow. The engineering team must have a thorough knowledge on many different disciplines, a non-complete list is indicated in the figure. These competences should also include in depth knowledge of physiological measurements and signals both for normal conditions but also for severe clinical conditions. From the user aspect knowledge of the clinical workflow and device use and user interface are important. Finally, the patient side and patient comfort are important aspects. These topics are part of this course. The measurements and extracted parameters of a typical parameter are discussed in the following section.

A short summary of the essential physics and physiology needed in this course and discussed subsequent lectures in this course is described in the next session. The education, approach and way of working in the healthcare domain is discussed thereafter.

2.1 Function of Monitor and Therapeutic devices

A patient monitor is a device or medical test for observing a biological condition, medical parameter or function of a patient over time. Monitoring provides essential information on the health status of a patient. The output data of a monitor (parameters, waveforms, trends, see Figure 2-10) is used for clinical decision making and for selection and optimization of therapy by a clinician. Note that monitoring is only useful when combined with therapy. The clinicians view is nicely illustrated by the following quote from Dr. Michael Pinsky: "*no monitoring device, no matter how simple or sophisticated, will improve patients' outcome unless coupled to a treatment which itself improves outcome*".

A patient monitor is a complex medical device which should be highly reliable, safe for the user and patient and compatible with the workflow of the clinicians. The main functions are listed below:

- Provide actionable data and information to the clinician such that best therapy can be administered
- Scientific evidence that outcome improves by (new) monitoring

The target group are clinicians, hospitals. It is important to realize that this is a different environment than the technical and engineering environment, with a different culture, way of working and technical expertise level. This will be discussed later.

Patient monitors and therapeutic devices have a physical contact with a patient. As mentioned before the monitor measures and monitors vital signs. In the terminology of industrial process analysis, monitors are **sensors** that provide **health care professionals** with **patient-related data and information** that supports those professionals in making **appropriate therapeutic decisions**. Therapeutic devices, in this terminology, are the actuators to perform therapeutic decisions.

There are several types of therapeutic devices, the two most important ones are listed below:

- Therapeutic devices to *support or replace* impaired or failing organs (e.g., mechanical ventilators, defibrillators, renal replacement therapy, may include monitoring devices).
- Therapeutic devices to *administer medications* and/or gases/fluids to the patient.

In this course the emphasis is on the therapeutic support devices.

Patient monitoring can be viewed as a process. This is illustrated in Figure 2-2. First the measurements provide data and waveforms of parameters that are of interest for the treatment. They are displayed on a screen (see Figure 2-3) in such a way that the information can be interpreted quickly by a clinician, this is the user interface. Based on the monitor data (and possibly other tests) a

decision is made on the choice of therapy for the next phase. When needed the final step is the adjustment of therapy. Note that this a continuous process, it goes on until monitoring is not needed anymore.

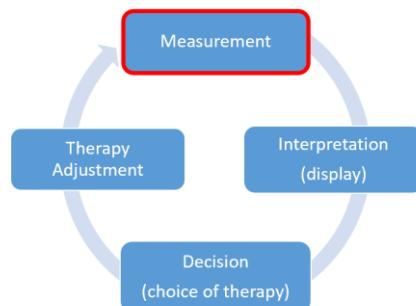


Figure 2-2 The main steps in the process of patient monitoring.

A mentioned before the focus of the clinical measurements is on the cardiovascular system and respiratory system added with a few other relevant parameters such as temperature. The main parameters or vital signs are illustrated in Figure 2-3.

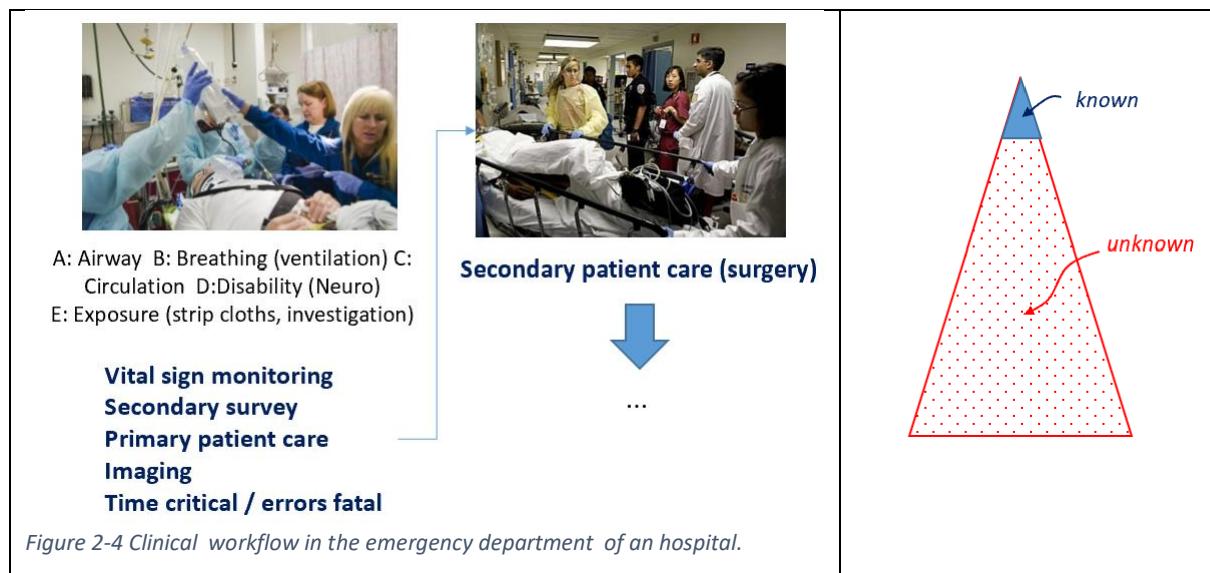


Figure 2-3 A patient monitor with the main measurements and vital sign parameters and user interface of a patient monitor.

The main vital signs are heart rate (HR, beats per minute), respiration rate (RR, breaths per minute), blood pressure, SpO_2 i.e. oxygen saturation of hemoglobin in arterial blood and temperature (fever). The display of the waveforms is of importance to give trust to extracted parameters but often also contain important information on the health status of the patient. For instance the electrocardiogram waveform has important information on the electrical system of the heart. For instance some deviations in the waveform indicate that the patient has suffered a myocardial infarction, even the location of the infarct can be determined. The main priority of patient monitoring is to maintain homeostasis, i.e. the steady state of the body that is compatible with life. Thus priority one is to maintain tissue perfusion and oxygen delivery, transport nutrients and other substances to the tissue and remove waste products such as carbon dioxide. A second function is to warn the clinician when parameters are out of the acceptable range, i.e. generate an alarm or warning of imminent danger. Note that the data displayed on the screen must be accurate and precise, almost real time (0.25 seconds latency), and that the waveforms must be close to the physiologic signals. Note that signals from the sensors often have low signal-to-noise, are prone to artifacts and can have a large dynamic range (in some cases $\text{SNR} > 100\text{dB}$.).

2.2 Differences between Engineering and Medical fields

There are large differences in the way of working, culture, organization, education, semantics and problem solving between the engineering- and hard sciences versus the medical sciences. This is illustrated with the help of an example of the workflow in the emergency care room when a trauma patient enters the emergency department (see Figure 2-4). The receiving team has little information on the injuries of the patient, they only know that patient has the highest injury score. There are many uncertainties facing the team and the situation can be very complex and hectic. The first action is to identify the main injuries and try to stabilize the patient and apply a list based so-called ABCDE protocol which specifies a predetermined chain of actions (protocol): the motto is treat first what kills first. Note the ABCDE mnemonic has direct meaning and helps the clinician in the list-based actions. The first focus is on the airway A (is it open) and breathing action (B), treat if needed. The second priority is the circulation of blood (C), are heart rate and blood pressure in an acceptable range? If not apply emergency therapy. With the ABC steps the clinicians ensure that oxygenated blood is transported to the tissues, a minimal flow of oxygenated blood to the brain and heart is the first priority and buys time for further treatment of the actual injury. The next steps focus on neurologic status (Disability D, brain, spine damage) and further examination after removal of clothing (E: external investigation, additional undetected injuries?). During the ABCDE procedure a vital sign monitor is connected to the patient, thereafter quantitative information on the main respiration and circulation parameters is available (although limited information). Note that the patient status can change rapidly and therefore measurement data must be monitored continuously. The whole process is extremely time critical and errors can be fatal. The clinicians must have the relevant medical knowledge immediately available and apply it at the same time (there is no time to search for knowledge that is not available). Besides following the protocol clinicians can use their experience and acquired pattern recognition to improve treatment quality for the specific patient. Although there is a scientific basis for the treatment steps there is still a lot unknown and the scientific base is less firm than in hard sciences. In short, the situation is hectic, there is a lot of uncertainty, the complexity is overwhelming and the time pressure is immense.



The medical education prepares the clinician for a wide range of medical conditions and subsequent diagnosis setting and therapies. During education medical knowledge and therapy steps are memorized. Diagnosis and treatment are list based processes that are memorized, they need not necessarily be understood. These lists change continuously with the progress of medical sciences.

During the medical education the focus is on memorization, know what to do in specific cases, how to handle (difficult) patients. Hence there is more attention on the human side of the profession than in the engineering sciences. There is little time spent on the “hard sciences” like physics, chemistry, physiology and computer sciences. Words and phrases can have a different meaning than in the engineering domain. The way of working and team organization and hierarchy is different.

For engineers and scientists active in the healthcare domain it is important to be aware of differences in culture, workflow, type of education and user needs and expectations. A brief review of the main differences between engineering and health care professionals is shown in Table 1.

Table 1: List of differences in education and approaches between engineering and medical professionals.

	Engineering	Medical
Way of working	Understanding, analytical	More memorization, protocols (complexity, unknown or obsolete knowledge)
Logic	Very important	Important (intuition & experience also)
Failures	More tolerated	Disaster
Information	More focused, know where to find it, well established	Sheer amount of data and rapidly growing, instant availability (emergency) Partly outdated
Science	Science and engineering practices well established System know how, scientific knowledge mature, less uncertainty, reproducible	Science (studies, statistics, physiology) <i>Evidence based methodology</i> <i>Protocols</i> differential diagnosis, therapy Time critical

Monitoring devices are often used during diagnosis setting. The medical diagnosis process is complex and challenging. The main steps are discussed in the following section.

2.3 Medical Decision Making and Diagnosis Process

Medical diagnosis is the process of determining which disease or condition explains a person's symptoms and signs. A symptom is subjective evidence of a disease like fatigue, pain. Only the patient can perceive the symptom and there can be many causes. A sign is objective evidence like a wound, a broken arm, a blood pressure. The diagnosis process is illustrated in Figure 2-5.

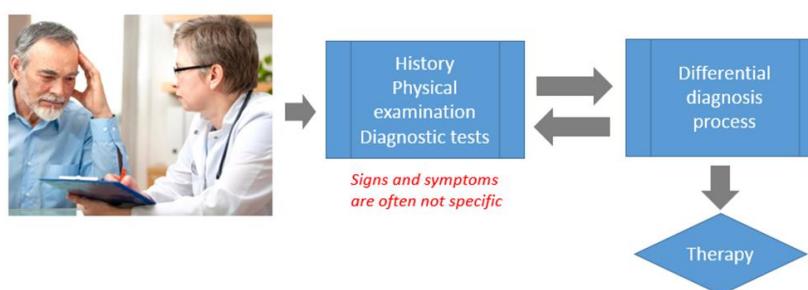


Figure 2-5 Illustration of the different steps in a diagnosis process.

Clinicians use a list-based process for finding the correct diagnosis given certain patient input. The doctor asks the patient for the reason of visit and for specific complaints, symptoms and signs. The doctor also investigates the medical history of the patient and does a physical examination and decides if additional diagnostic tests are needed as signs and symptoms are often not specific. The information obtained so far is then used in a differential diagnosis process. Differential diagnosis is the distinguishing of a particular disease or condition from others that present similar clinical features. This process can lead to further questions and diagnostic tests. It is an example of a classification process. It is a list-based process (memorization) and is an elimination of different options based on available probabilities (often only rough estimates). The basis is Bayes theorem (use of prior knowledge). It is often a complex process given the large uncertainties in input data and knowledge, data quality, comorbidities (other diseases present) and poor description of symptoms. It works best for average patients and the process is well suited for computer assisted solutions and support. Note that in general the patient is not an average patient. An example of a differential diagnosis and subsequent selection of therapy is given below in xxx. A patient with a medical history in cardiac disease suffers from symptoms that indicate possible cardiac arrhythmia (irregular heartbeats). The main steps are shown in the table.

The list based differential diagnosis process led to overwhelming evidence that atrial fibrillation was the cause of the symptoms and signs. Based on this diagnosis the doctor decided to start a therapy based on medicines that suppress the irregular heart rhythms and anti-thrombolytic medicines to suppress blood clotting. An appointment with the patient is made in two months to verify if the therapy is successful.

Table 2 Steps in a differential diagnosis process for cardiac arrhythmia symptoms.

<i>Actions (in prescribed order)</i>	<i>Result</i>
Gathers all information about the patient and creates a symptoms list.	Medical history (heart failure), fatigue, anxiety irregular heartbeat, shortness of breath, dizziness, leg pain
Lists all possible causes (candidate conditions) for the symptoms	Atrial tachycardia, ventricular tachycardia, Wolf-Parkinson-White Syndrome, atrial flutter, atrial fibrillation (AF), blood clotting process
Prioritizes the list by placing the most urgently dangerous possible causes at the top of the list	Sorts the list and decides for a 12 lead ECG diagnostic test (electrical activity of the heart)
Rules out or treats possible causes, beginning with the most urgently dangerous condition and working down the list	 ECG indicates that atrial fibrillation (AF) is present, other causes ruled out or less likely
Additional diagnostic tests (verification of diagnosis and find cause of leg pain)	Chest X-ray, ultra sound Imaging (heart, legs), MRI scans, blood tests. Found that blood clot in the femoral artery causes ischemia and leg pain (related to AF). All tests confirm atrial fibrillation as most likely diagnosis.

2.4 Modules of the course

The course is split into 5 modules, the first is an introduction to the course that is described in this chapter of the reader, the second is a short recapitulation of the essential physics that is needed to master the material on sensors and clinical measurements. The third module is on the most common sensors and transducers that are used in patient monitors. The forth module discusses the main clinical measurements of the cardiovascular and respiratory systems. In the fifth module integrated devices, i.e. patient monitors and therapeutic devices are discussed. Below these modules are briefly described.

2.4.1 Module 2: Physics and Physiology Topics in the Course

A basic background knowledge of physics of elasticity and fluid dynamics, physiology and pathophysiology are needed for the remainder of this course. This knowledge is needed for the sensing (how) and the clinical measurement (what is measured) parts. There is a basic introduction in essential physics of fluid flow and elasticity in the oral part of the course. The physiology part is covered in the practical part of the course by two pre-recorded lectures on physiology (optional for students who have followed bio-medical courses) and in the practical simulation exercises. Simple lumped element models are used throughout this course for both physiology and sensing systems. These simplified first and second order models are important for understanding the complex signals and systems that are part of the course. Note that these lectures are also of importance to interpret the signals and parameters that are measured as well as to provide knowledge of acceptable ranges of these parameters.

The most important parameters shown on the monitor are related to pressures and flows in the cardio-vascular and respiratory systems. Understanding of the physiology of circulation and respiration helps the clinician for interpretation of the data. As an example Figure 2-6 shows flow of blood from the supply side (high pressure artery) to the return site (veins) via ever smaller diameter arterioles to the capillaries where exchange of nutrients, gases and (waste) products takes place. Note that mean flow is in the direction from high to low pressure. The arterial-venous blood pressure difference must be sufficiently large to guarantee minimal flows to maintain cell metabolism. Therefore, this flow on a micro circuit level is a crucial parameter for the clinician. It is however very difficult to measure parameters in the micro circulation. Other surrogate parameters related to perfusion in the microcirculation are more practical and widely used in clinical practice.

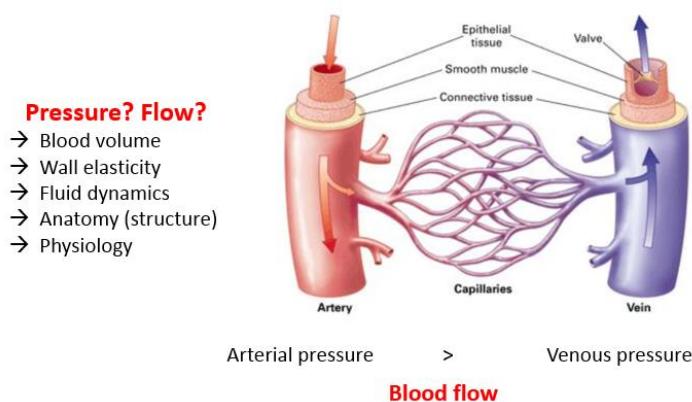


Figure 2-6 Schematic drawing of blood flow and pressures in the vascular system.

The mechanical properties of the elastic walls of the blood vessels and bronchi in the lungs are of paramount importance for correct physiologic functionality. Abnormal wall elasticity is linked to severe diseases (e.g. hypotension, arteriosclerosis, myocardial infarction, stroke, lung diseases like

asthma, emphysema etc.). In module 2 a short introduction in basic elasticity of (bio-) materials is presented. Pressure gradients drive flow. The science of the flow of fluids due to external forces is fluid dynamics, one of the most complicated branches of physics. The dominant types of fluid flow in blood vessels and the airway are presented in module 2. The anatomy (structure) and physiology of the circulation and respiration systems are presented in prerecorded lectures 1 and 2 of the practical part of the course.

2.4.2 Module 3: Sensors Transducers and Clinical Measurements

The sensors, transducers and measurements that are typically used in a patient monitor are shown schematically in Figure 2-7. The measurements can be invasive (i.e. sensing occurs inside the body) or non-invasive (sensors located outside the body). Invasive measurements are obtrusive and provide higher quality continuous waveform and parametric data (e.g. blood pressure) but pose a higher risk to the patient (e.g. puncturing holes in high pressure arteries, infection). They are used mostly when the need for high quality and continuous data outweighs the higher risk for the patient. The most common sensors and the definition and basic architecture of clinical measurements are described in module 3.



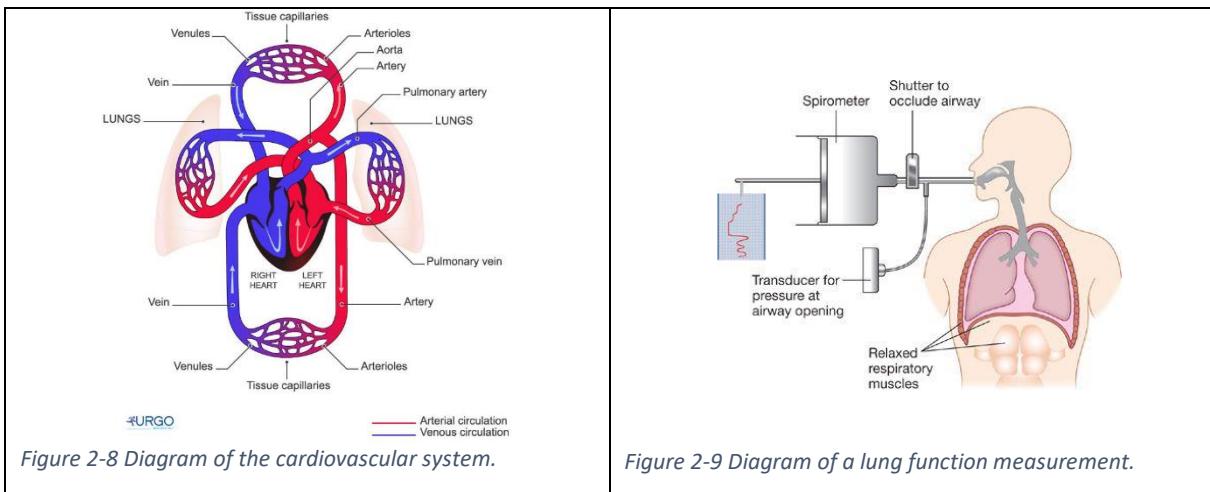
Figure 2-7 Sensors and measurements of a patient monitor.

2.4.3 Module 4: Measurements of the cardiovascular and respiratory systems

In this module the main measurements of the cardiovascular and respiratory systems are discussed. Diagrams of the cardiovascular and respiratory systems are shown in Figure 2-8 and Figure 2-9.

Measurements of the cardio-vascular system include electrical and mechanical properties of the pump, i.e. the heart, the blood flow out of the heart (cardiac output), flow in large arteries, pulse propagation in arteries and blood pressures in the large arteries and oxygenation of blood in the peripheral small arteries. Both invasive and non-invasive measurements are discussed.

Measurements of the respiratory system include airflow into and out of the lung, pressures in the airways and lungs, flow resistance of the airways, lung volumes, elasticity of the lung (compliance) composition of inspired and expired air and test of the lung function. After module 4 the sensors, and main clinical measurements have been discussed and the integration of these measurements into a patient monitor device is discussed. This is the topic of module 5.



2.4.4 Module 5: Patient Monitors and Therapeutic devices

In module 5 state of the patient monitors and therapeutic devices are discussed. A high-end patient monitor device and a rack with clinical measurements is shown in Figure 2-10. This type of monitor is typically used at a high level of care (high-acuity), i.e. in the intensive care department or in the operating theatre. The system requirements are discussed first followed by a discussion of the architecture of a modern high-end patient monitor. Then the parameter and feature extraction is discussed. Thereafter methods for alarm generation are discussed. False alarms and alarm fatigue are major issues that are discussed in detail. Finally challenges and issues for this type of device are discussed.



Figure 2-10 A patient monitor with on screen content (waveforms and numeric data) and a module with measurements..

In the past patient monitoring devices were mostly used in the ICU and OR(i.e. high-acuity). There is a growing demand for monitoring of patients at lower levels of care (low acuity). Examples are a hospital ward and monitoring of patients of chronic diseases and patients discharged from the hospital to limit readmission to the hospital. This type of monitoring had different requirements because patients are mobile, instead of alarms warning are generated and devices are battery

operated and need wireless connectivity and device cost must be very low. These are major challenges and low acuity monitoring is considered to be the next frontier in patient monitoring. Finally all types of monitoring generate huge amounts of data that is underutilized. Data analysis and machine learning are widely studied to tackle this problem. The status and challenges in these fields are discussed in the last lecture.

In module 5 therapeutic devices are described, the emphasis is on mechanical ventilation devices. An advanced device and its breathing set (the tubes toward the airway opening of the patient) are shown in Figure 2-11. Mechanical ventilation is a live saving technique but optimization of ventilation parameters for a patient is difficult, especially as the patient lung condition can vary appreciably during treatment. The basics of a mechanical ventilators is discussed, different modes of operation are explained and the impact of ventilation parameters such as pressure, flow and volume on patient injury is discussed.

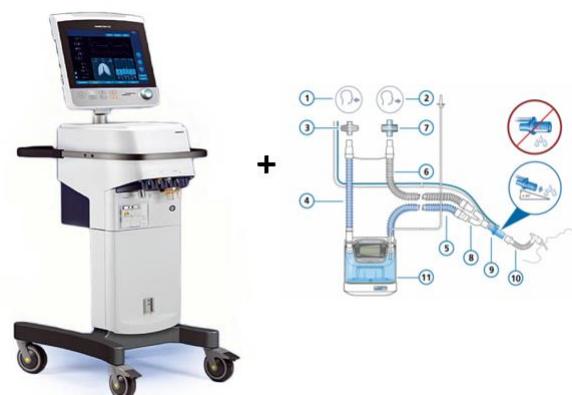


Figure 2-11 High-end mechanical ventilator and breathing set.

In the final part of module 5 the use of monitors and therapeutic devices during time critical care and low acuity care is discussed. The module ends with a discussion of the outlook and the main future technologies both in hardware and data analysis.

2.5 Questions

1. Describe the process through which a clinician reaches a diagnosis.
2. What is the function of a monitor in the clinical care process?
3. Describe the main steps in the process of patient monitoring.
4. What is the function of a therapeutic device?
5. Describe how a patient monitor is integrated in the hospital system.
6. What are the main differences between engineering and clinical education?

3 Elasticity and Viscoelasticity

Blood and air are transported in the human body via elastic tubes. The heart and lungs function as the pumps that generate forces, pressures and flows in a system of elastic tubes. The elastic properties of these tubes, tissues and organs are of paramount importance for the functioning of the circulation and respiration body systems and have a large impact on vital sign parameters like blood pressure and blood flow. Furthermore elasticity is important in many sensing systems. The basic physics of elasticity that is needed for the remainder of the course is discussed in this chapter. The tube walls and fluids exhibit both elastic and viscous properties (damping, energy dissipation) and its mass (inertance) can be important. The biomaterials in the body are viscoelastic.

An elementary introduction to the elasticity and viscoelasticity of biologic materials and simple cylindrical tubes is presented in this chapter. For readers that want to have more in depth information on these topics references to text books are added. Before discussion of elasticity of tissues, force and pressure are defined and the different units for pressure and flow that are used by clinicians are described.

3.1 Force and Pressure

Deformation of elastic materials is important for the physiology of respiration and circulation. When a force is exerted on an object its velocity changes. Think off a push or a pull on an object like a ball during soccer. When the object cannot move, it is static, and forces will deform the object. Think of a rubber band that is stretched. A force is a vector and has a magnitude as well as a direction. The direction and magnitude of a force determine how the object will move. The SI unit of force is kg m/s^2 in short Newton (N).

Pressure is related to the magnitude of a force; it is a scalar quantity. Pressure is defined as force per unit area. It is important to remember that pressure is a force density and that pressure is more important for physiology than force. Forces on a surface can be calculated once the pressure, contact area and its shape are known.

When a pressure acts on a plane of an object a force is exerted perpendicular to the surface. The orientation of the plane is not of importance, the magnitude of the force exerted is the same for each orientation, the direction of the force differs, the force is perpendicular to the surface. An example is shown in Figure 2-1. The pressure is on the top of the surface. The net force is perpendicular to the surface and is in the present case directed in the opposite direction of the unit vector \vec{n} as this unit vector points always in the outward direction of the plane. This is the origin of the minus sign in the equation. Since the pressure is uniform the force magnitude is simply the product of area and pressure.

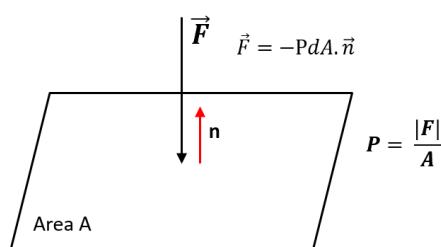


Figure 3-1 Definition of pressure. Pressure is force per unit area. The force is a vector that is in the opposite direction from the normal unit vector n that always points outwards.

For non-planar surfaces S like in Figure 3-2 the local force vector can be determined in a similar manner as in Figure 3-1 the force at area dA is the product of the scalar quantities of pressure P and area dA times the local unit vector \vec{n} . In this case the pressure on the inner surface is considered, the force points in the same direction as the unit vector. The total force exerted on the curvilinear surface S is the surface integral $F = \iint P \cdot \vec{n} dA$

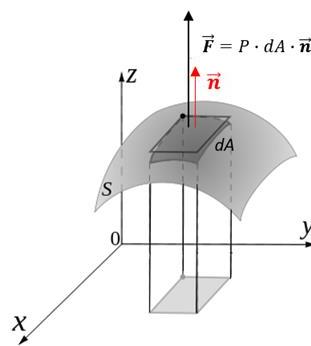
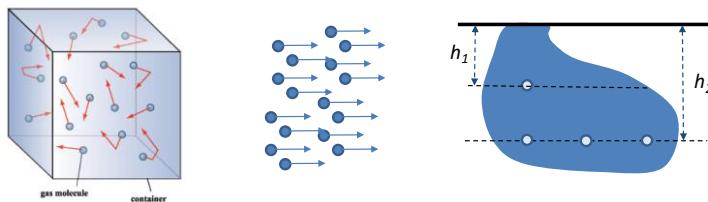


Figure 3-2 Relation between force and pressure for a curve linear surface.

There are three types of pressure. The internal pressure P , the dynamic pressure P_d and the hydrostatic pressure P_h . In Figure 3-3 the different types are shown in a schematic drawing.

Internal Pressure Dynamic Pressure Hydrostatic Pressure



$$P \quad P_d = \frac{1}{2} \rho v^2 \quad P_h = \rho g h$$

Figure 3-3 Illustration of the three different types of liquid pressure.

In a fluid the molecules move around in random directions due to thermal motion. The internal pressure (P_i) of the fluid is caused by momentum transfer due to collisions of molecules with the container walls. When there are other forces on the fluid (like gravity) the internal pressure changes. **The internal pressure is the most relevant pressure for clinicians.**

When there is a **net flow of volume or mass in the system** a drift velocity v is superimposed on the random thermal motion of the individual particles. The kinetic energy of the liquid with net flow velocity v causes a **dynamic pressure** (P_d) that is equal to $\frac{1}{2} \rho v^2$ where ρ is the fluid density.

The third type of pressure occurs in a fluid when a **gravitational force** is present, it is called the **hydrostatic pressure** (P_h). The gravitational force due to the liquid mass above a certain coordinate causes a hydrostatic pressure component. The local additional hydrostatic pressure component (P_h) depends on the position in the fluid and is equal to $\rho g h$ where g is the earth gravitational acceleration and h is the position with respect to a reference level, often the highest position, for instance the surface of the fluid-air interface. In cardio-vascular system the reference position is in the atrium of the right heart. Note that the hydrostatic pressure depends only on depth (see **Error! Reference source not found.**), the hydrostatic pressure at position h_1 is lower than at position h_2 . **The pressure at the three points at h_2 are identical as only the height matters, the geometry is irrelevant (Law of Pascal).**

Pressure can be measured with respect to a reference pressure. When a **zero-pressure reference** (vacuum) is used, it is called the **absolute pressure**. In most cases the reference pressure is the atmospheric pressure and this pressure difference is more relevant in the clinical domain. A

convenient standard pressure that is often used is the atmospheric air pressure at sea level at standard conditions. Pressures measured with respect to atmospheric pressure are called gauge pressure. Pressure differences between the inner and outer surfaces of a tube are transmural pressures. In this book the gauge or transmural pressure is used most frequently. As noted before pressure differences are the most relevant ones in the physiology of circulation and respiration.

The official SI unit of pressure is N/m² or Pascal (Pa). The Pascal can also be expressed in different units such as kg /ms² or Joule/m³. Hence pressure is an energy per unit volume. The standard atmospheric air pressure is another unit that is often used. It is equal to 101325 Pa. There are many other units, in the medical field two other units are still in wide use, most of the clinicians are not using SI units.

In the 17th century Torricelli invented the mercury manometer to measure both absolute and gauge pressure. The difference in liquid height between the inlet and reference tube is proportional to the height difference between the two mercury surfaces, when the pressure difference is one atmosphere the height is 760 mm. Mercury was chosen because it is the liquid with the highest density, more than 13 times larger than that of water. For measuring pressures in the order of a tenth of one atmosphere mercury is a very practical choice as the column height is in the order of tens of millimeters. When water would have been chosen the height of the liquid column would be more than a meter. This is unpractical for larger pressures. Before the invention of electronic pressure sensors these manometers were the main devices for measuring pressures with high accuracy and precision. They were widely used by clinicians.

For the circulatory system gauge pressures of blood are between 0.1 and 0.2 atmospheres and therefore mercury manometers were the system of choice. These devices have been used for more than a hundred years but are now replaced because of the toxicity of mercury. Clinicians are still using the unit of mmHg for the pressure in circulatory measurements. For the respiratory system the air gauge pressures are in the order of 0.01 atmosphere. Hence for measurements in the respiratory system water was more practical than mercury and water manometers were widely used during lung measurements. Most clinicians still express airway pressures in cmH₂O. The use of SI units is still not in widespread use in clinical practice. Therefore, in this course book ‘medical’ units are used, the value in SI units is shown between brackets when relevant. Organs, bronchi and blood vessels deform when forces and pressures are exerted on these bodies. Furthermore elasticity of tissues is very important for many physiologic processes. The basic physics for (elastic) deformation is discussed in the next section.

3.2 Elasticity and Material Deformation

When a force is exerted on a static object, its shape changes (see Figure 3-4). For an ideal elastic material the shape of the object returns to the original shape when the force is removed. For small forces the relation between length change and force is linear. This behavior is similar to the behavior of an ideal spring. For an ideal spring the elongation ΔL of the spring is linear proportional to the applied external force F (see Figure 3-5). The proportionality constant k is the spring constant, the larger k the stiffer the spring. This is known as Hooke’s law. This equation has a validity beyond that of a spring it is also valid for deformation of solid objects. It also holds for small deformation of objects due to torques T (product of force and distance to point of attachment applied in tangential direction) and shear forces (force F applied parallel to a surface, see Figure 3-6).

When forces are very large the relation between force and length becomes non-linear and for larger changes in length the shape can change permanently, this is called plastic (or permanent)

deformation. When forces are increased even further the force can exceed the material strength and fracture can occur.

In this course elastic properties of biomaterials are described. Note that these are very complex materials with properties in between an elastic solid material and a viscous fluid, they exhibit so-called viscoelasticity. First the deformation of ideal elastic material is discussed followed by an introduction into viscoelasticity and the viscoelasticity of biomaterials. Thereafter inertance (mass of the object) is added and impact on equations of motion is discussed.

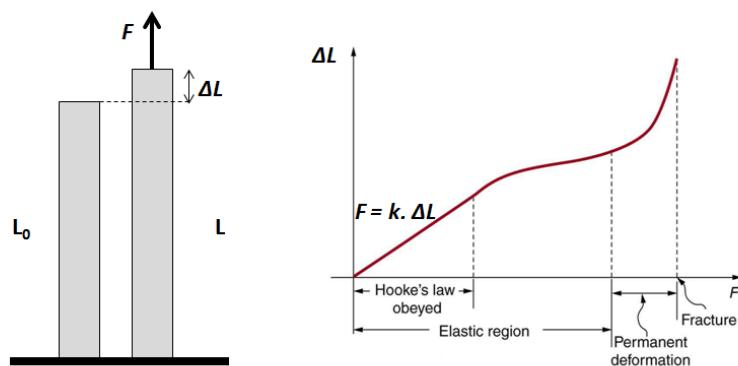
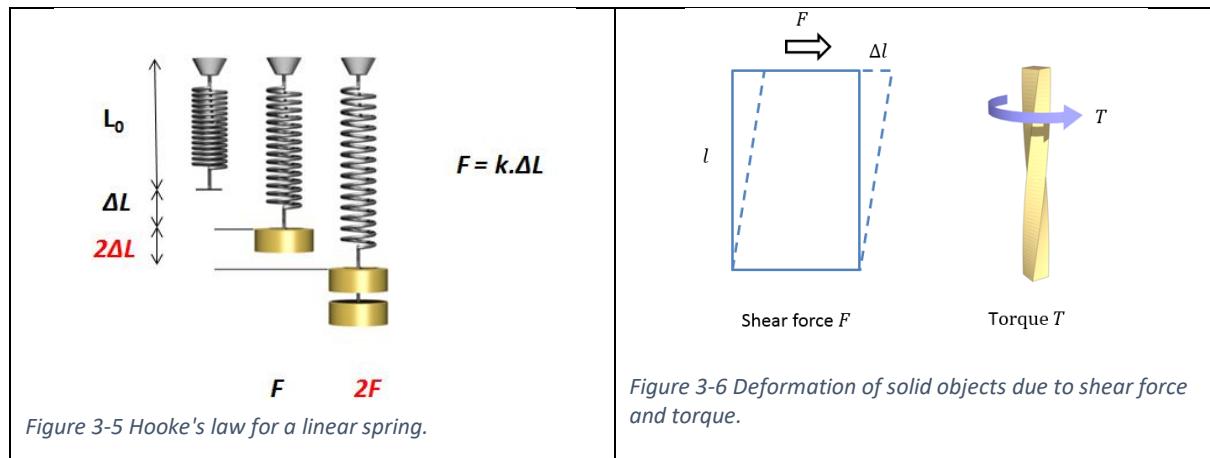


Figure 3-4 Change in length L of a bar due to an applied force F . On the right a graph of length change ΔL as function of the magnitude of the applied force. Different regions of deformation are indicated.



3.2.1 Elastic Deformation of materials

Deformation of an object due to forces depends on its shape and material properties. As an example consider the elongation or compression of an ideal homogeneous elastic material with rectangular shape with length l , side length a and cross-sectional area A (see Figure 3-7). The elongation or compression due to the applied force is referred to as Δl and is equal to: $(l - l_0)$ with l_0 the length of the bar without force applied. For a tensile force Δl is positive, for a compression force Δl is negative. For both forces there is also a change in the direction perpendicular to the force, this is denoted by Δa , Δa is positive for compression and negative for tension.

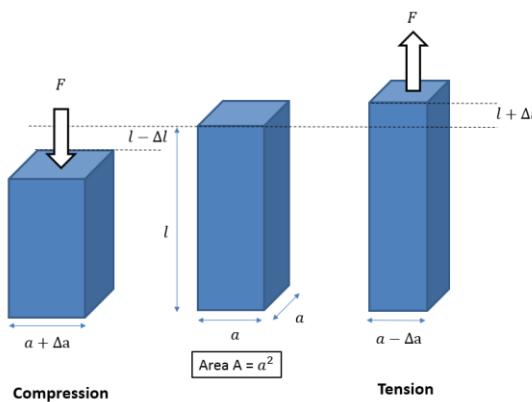


Figure 3-7 Change in dimensions due to compressive and tension forces.

It is found by experiments that the length change is proportional to the applied force F , the proportionality constant has a positive sign:

$$\Delta l \propto F$$

The change in length is also proportional to the length l_0 in the direction of the force F and inversely proportional to the area A :

$$\Delta l \propto l_0 \text{ and } \Delta l \propto 1/A$$

This results in the following relation:

$$\Delta l \propto \frac{F}{A} \cdot l_0 \text{ or } \frac{\Delta l}{l_0} \propto \frac{F}{A}$$

This can be written as:

	$\frac{F}{A} = E \frac{\Delta l}{l_0}$	3.2-1
--	--	-------

The proportionality constant E is a material property and is called modulus of elasticity or Young modulus, E has the dimensions of force per unit area, in SI units N/m^2 or Pascal (Pa). For small force and elongation, the elastic modulus E is constant, for large forces and elongations E will depend on these parameters. The larger E , the more force is required for a certain elongation, in other words the larger E the stiffer the material (e.g. steel versus rubber). A small E belongs to a material that can easily be stretched (e.g. rubber). Values for the Young Modulus E are listed in Table 3 for some engineering and biomaterials. In the mechanical engineering domain, a different notation is often used. The relative length change $\varepsilon = \frac{\Delta l}{l_0}$ is called strain. The force per unit area $\sigma = \frac{F}{A}$ is called stress.

Equation 3.2-1 can be expressed in the parameters stress and strain as:

	$\sigma = E\varepsilon$	3.2-2
--	-------------------------	-------

Note the elastic modulus E of engineering materials is very large with the exception for some rubber materials. For biologic tissues the elastic modulus is much smaller than that of rubber and the stress-strain relation is often highly non-linear. The biomaterials are very elastic for small strain but become very stiff for larger tensile strain (e.g. collagen is such a material). Furthermore, the range of E for tissues is very large from 100 Pa (lung) to 10 GPa (bone).

Table 3: Elastic modulus of common engineering materials and biological tissues.

Material	Young Modulus E (GPa)
rubber	0.01 – 1
wood	10
concrete	30
Steel	200
Diamond	1050
Elastin	$\sim 10^{-4}$
Collagen	10^{-4} - 1
Fat	$\sim 10^{-6}$
Muscle	$\sim 10^{-5}$
Lung tissue	$\sim 10^{-7}$
Bone	10

For an ideal elastic material, the energy supplied to deform the object is stored in the material in the form of elastic energy. This is a form of potential energy. This elastic energy is released when the body returns to its original shape, i.e. when the applied stress is removed. There is no loss in energy with loading and unloading, the elastic energy is conserved. The elastic potential energy U_{el} (unit Joule) for an applied length change ΔL is equal to:

$$U_{el} = \int_0^{\Delta L} F(x) dx = \int_0^{\Delta L} \frac{E \cdot A}{L_0} \cdot x dx = \frac{E \cdot A}{2L_0} \Delta L^2 = \frac{1}{2} K \Delta L^2 \quad 3.2-3$$

Here K is the elastic constant of a bar with length L_0 , cross sectional area A and Young Modulus E . The elastic energy is proportional to the product of K and the square of the length change ΔL . The elastic energy E_{el} per unit volume (units Joule/m³, divide U_{el} by $A \cdot L_0$) is equal to:

$$E_{el} = \frac{1}{2} E \varepsilon^2 \quad 3.2-4$$

The potential energy stored in the elastic material is very important in both respiration and cardiovascular physiology. This concludes the deformation and energy storage in ideal solid elastic materials. In physiology the relation between volume, radius and pressure of hollow thin walled elastic structures (blood vessels, lungs, bladder, heart, and stomach) is of great importance. This is the topic of the next section.

3.2.2 The Laplace Law in Elasticity

In the section below the first order relation between dimensions and pressures for thin walled hollow elastic structures is discussed. We want to calculate the pressure-volume relation of a thin-walled cylindrical tube with inner radius r and length ℓ (see Figure 3-8). When the fluid pressure inside (P_{int}) the tube is larger than the outside pressure (P_{ext}) the elastic wall is stretched (strained in the circumferential or tangential direction) and a circumferential stress in the tube wall is generated (i.e. wall stress). The transmural pressure P_{tm} is defined as: $P_{tm} = P_{int} - P_{ext}$. For a positive transmural pressure (internal overpressure) the radius increases from r_0 to r (see Figure 3-8). In equilibrium the net force on the cylinder wall is zero, the force generated by the transmural pressure is balanced by the elastic wall stress.

For the derivation of the Laplace equation we divide the cylinder in two halves and calculate the force components perpendicular to the mid plane (see Figure 3-9). The positive transmural pressure

induces a net vertical force¹ F_p on the upper half of the cylinder. This net force is perpendicular to the axial plane between the two halves. This force is balanced by two elastic forces F_e at the left and right sides induced by the elastic pull of the lower half of the cylinder. In equilibrium the forces must balance, this gives:

	$F_p = 2F_e$	3.2-5
--	--------------	-------

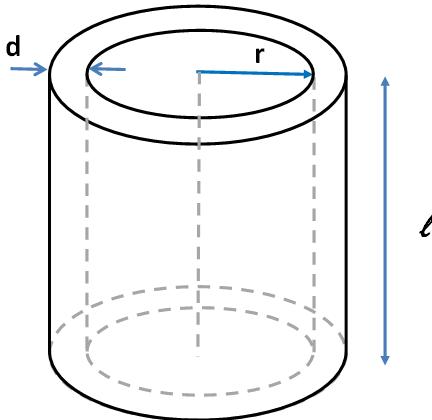


Figure 3-8 Dimensions of a thin walled cylinder with ideal elastic wall.

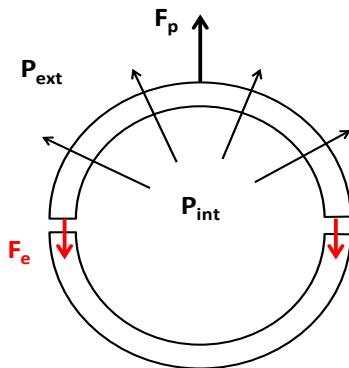


Figure 3-9 Forces (F) caused by the transmural pressure (F_p) and the elastic wall stress (F_e).

The force F_p is equal to the product of the transmural pressure P and the inner area of the axial plane (i.e. $2r \cdot l$). Hence:

	$F_p = 2Prl$	3.2-6
--	--------------	-------

The elastic force F_e can be obtained from the wall stress and wall area of the cross section:

	$F_e = \sigma_w \cdot d \cdot l$	3.2-7
--	----------------------------------	-------

Combining equations 3.2-5, 3.2-6 and 3.2-7 gives:

	$\sigma_w = \frac{P \cdot r}{d}$	3.2-8
--	----------------------------------	-------

This is a very important equation; it states that for a given transmural pressure, the wall stress σ_w is proportional to the radius r and inversely proportional to the wall thickness d . Hence at a given transmural pressure for larger r the wall tension is proportionally larger. The wall stress is proportional to the transmural pressure, the proportionality constant is the ratio of the radius and wall thickness. This is very important for physiology and pathophysiology; it will come back in later parts of the course.

This equation resembles the equation that Laplace derived for soap bubbles and liquids in contact with other liquids or solid materials. For bubbles the thickness of the wall is very small and wall tension T instead of wall stress is used. Wall tension is equal to $T = F_e/l$ it is the elastic force per unit length. Rewriting equation 3.2-8 in terms of wall tension T yields:

	$T = P \cdot r$	3.2-9
--	-----------------	-------

In the field of physiology equations 3.2-8 and 3.2-9 are referred to as Laplace's equation. Equation 3.2-8 is used in the cardiovascular domain and equation 3.2-9 is used in the pulmonary domain. For a

¹ The horizontal components of the force cancel out

sphere the derivation of the relation between wall tension and pressure P can be derived in a similar manner, the result is:

$$T = P \cdot r/2 \quad 3.2-10$$

For a sphere the wall tension T is only half of that of a cylinder for the same transmural pressure P . For an arbitrary curved surface, the local tensions and curvatures in two perpendicular tangential directions 1 and 2 are related to P by:

$$P = \frac{T_1}{r_1} + \frac{T_2}{r_2} \quad 3.2-11$$

3.2.3 Volume and compliance of a thin-walled elastic tube

In appendix 12.1 a first order equation for the volume-pressure relation of an ideal thin walled cylindrical elastic tube is derived. The volume-pressure relation is:

$$V(P) = \frac{V_0}{\left(1 - \frac{Pr_0}{Ed}\right)^2} \quad 3.2-12$$

Note that the relation between volume and transmural pressure P is **non-linear**, this is typical for elastic structures. The volume V_0 and radius r_0 of the tube are the unstressed (i.e. at zero transmural pressure) values. The elastic modulus E is a constant and it is assumed that both the Laplace and Hooke relations are valid. A very important parameter of an elastic structure is the differential compliance C at pressure P which is defined as:

$$C(P_1) = \left(\frac{dV}{dP}\right)_{P_1} \quad 3.2-13$$

The compliance is a measure of the volume change per unit pressure change, it is the tangent of the volume-pressure curve at volume V_1 and pressure P_1 . Note that the compliance depends on pressure. In physiology this parameter is called dynamic or differential compliance. This is a very important parameter for cardiovascular and pulmonary physiology and for sensors. It is discussed in more detail in the next chapter on fluid dynamics.

In physiology in some cases the static or large signal compliance C_s is used, it is defined by:

$$C_s = \frac{V_1}{P_1} \quad 3.2-14$$

This value can differ a lot from the differential compliance. Note that when the volume-pressure relation is linear the compliance is a constant and both dynamic and static compliances are identical. In calculations the differential compliance must be used.

3.3 Biologic tissues and viscoelastic deformation of materials

Biologic tissues are complicated materials. The elastic stress-strain curve is not linear and has a typical shape as shown in [Figure 3-10](#). For small strain the material has a linear stress-strain relation and the elasticity modulus E is relatively small, the material is relatively compliant. For larger strain a progressively larger stress is needed, the material becomes stiffer and the elastic modulus E increases. For the non-linear stress-strain curve the incremental elastic modulus E_{inc} is defined, the definition is given in equation 3.3-1 and in [Figure 3-10](#). It is the slope of the stress-strain curve at a certain strain. E_{inc} is an important parameter for mechanical tissue properties.

$$E_{inc} = \frac{d\sigma}{d\varepsilon} \quad 3.3-1$$

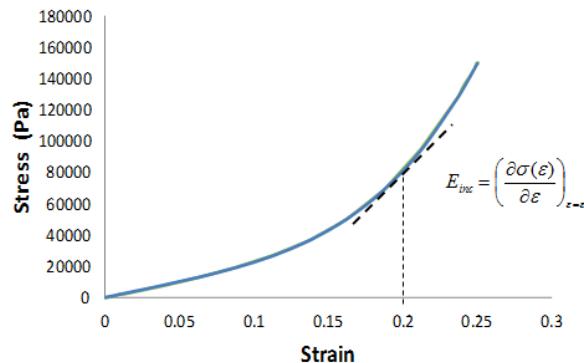


Figure 3-10 Stress-strain curve for a biologic tissue and definition of incremental elastic modulus.

For an elastic material the relationship between stress and strain is described by Hooke's Law (eq. 3.2-8Error! Reference source not found.). Hence for a pure elastic material the stress depends only on the strain and when a stress is applied the strain changes instantaneously, there is no delay, there is no time dependence. Ideal elastic materials² show time-independent behavior when a step force is applied, i.e. change in dimensions is instantaneous.

Biologic materials³ behave quite different, the stress depends both on the strain and on the rate of change of the strain $\dot{\epsilon} = \frac{\partial \epsilon}{\partial t}$, and the difference between dynamic loading between the two types of material is shown in [Figure 3-11](#) and [Figure 3-12](#). For an elastic material the stress-strain curve does not depend on $\dot{\epsilon}$, for a biomaterial the stress-strain curve is different for the loading and unloading trajectory, the difference increases with increasing $\dot{\epsilon}$. This is called hysteresis. The stress is now a function of both ϵ and $\dot{\epsilon}$.

$$\sigma = \sigma(\epsilon, \dot{\epsilon})$$

3.3-2

This time dependent mechanical material behavior is called viscoelasticity. The name consists of two material properties, viscosity and elasticity. Viscosity is a property of a fluid and is a measure of the resistance to flow by internal friction. The magnitude of the friction forces depends on the viscosity parameter η of the material and the rate of change. Elasticity is a material property of a solid.

Viscoelastic materials exhibit properties from both solids (elasticity) and liquids (viscosity).

Viscoelastic soft tissues consist for a large part of fluids (water > 60%), the rest consists of cells, fibers and other elastic materials.

A viscoelastic material has the following properties:

1. Hysteresis in the stress-strain curve. A different trajectory is followed in the stress-strain plot during loading and unloading (see [Figure 3-12](#)).
2. During loading and unloading energy is dissipated in the form of heat by internal friction. This is an irreversible process; energy is lost to the environment. The area or integral of the stress-strain loop in [Figure 3-12](#) is the energy dissipated during loading and unloading.
3. Stress relaxation, for a constant strain, the stress decreases gradually.

² Note that mass is neglected, i.e. an elastic structure without mass is assumed

³ and many other materials including polymers, plastics, fluids

4. Creep, for a constant applied stress the strain increases gradually.

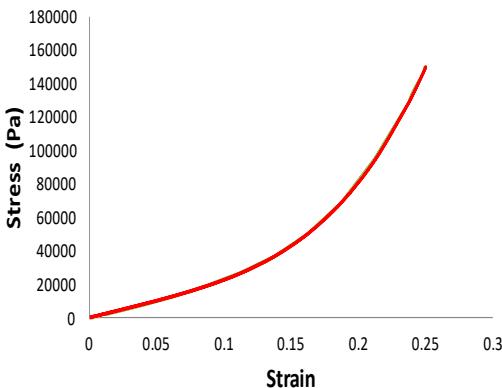


Figure 3-11: Stress-strain curve for an ideal elastic material for static and dynamic loading conditions. There is no time dependence.

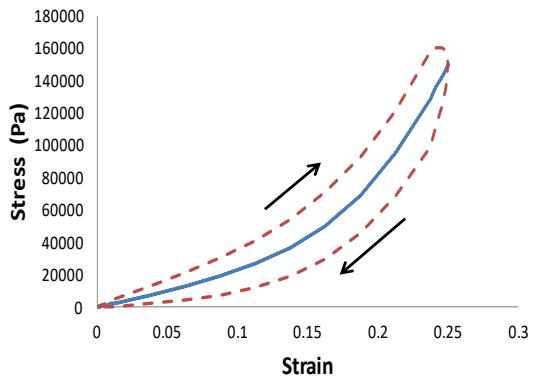


Figure 3-12: Stress-strain relations of a viscoelastic material for static (blue) and dynamic loading conditions (red dash). The dynamic case shows hysteresis.

Mechanical models for real viscoelastic materials are complex. For practical reasons the stress-strain relation and the temporal behavior of such materials are modeled using lumped element models formed of ideal springs and dampers. For the springs Hooke's law is used. The elastic lumped model is illustrated in Figure 3-13.

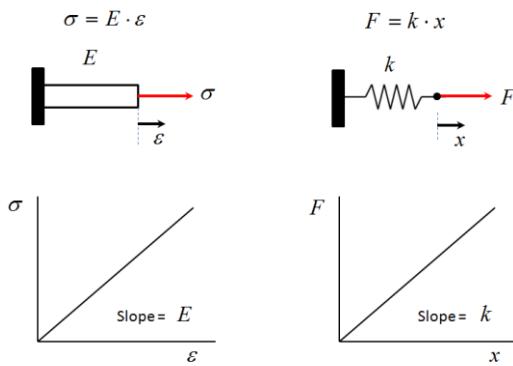


Figure 3-13 Analogy between elastic deformation of a material and a lumped model for the elastic deformation based on Hooke's law for a spring.

For the viscous part of the deformation the stress-strain rate dependence resembles that of a damper or dashpot. A dashpot is a device that is used to damp shocks and vibrations⁴, it resists motion by viscous friction. Dampers are used widely in the mechanical engineering field; a well-known example is in the wheel suspension of a car. A schematic drawing of a dashpot is shown in Figure 3-14. The device consists of a sealed cylinder filled with a liquid (typically oil is used but air can also be used). When the piston moves in the direction x with velocity $v = \frac{dx}{dt}$ a frictional force F_d arises which opposes motion. The frictional force is proportional to v and the proportionality constant is k_d , the damping coefficient (see Figure 3-15). The magnitude of k_d depends on the viscosity

⁴ An example is the damper in the suspension of a car

of the fluid and on engineering parameters. For the visco elastic material the main parameters for the stress are the viscous constant and the rate of change of the strain.

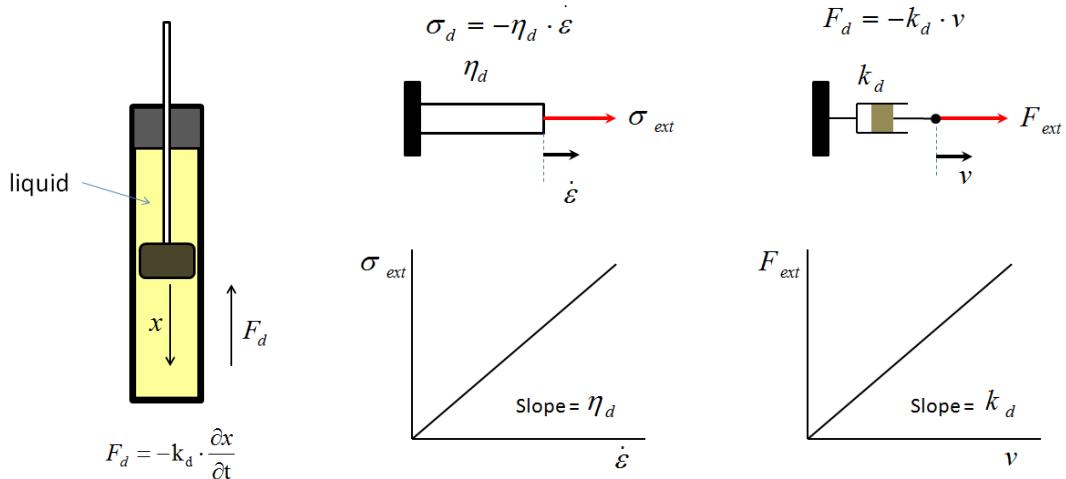


Figure 3-14 Schematic drawing of a dashpot.

Figure 3-15 Analogy between the deformation of a viscoelastic material and a dashpot. Note that the elastic deformation force has been assumed to be negligible.

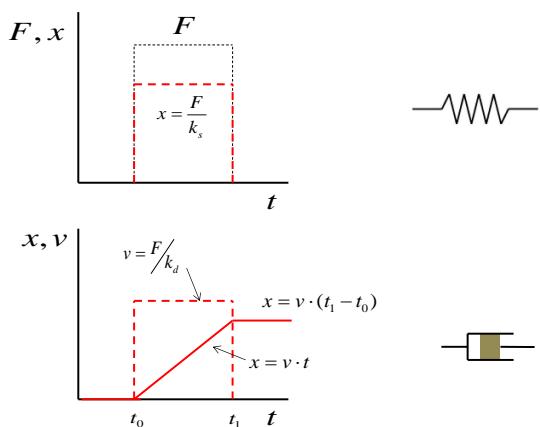


Figure 3-16 Time dependence of strain x for an ideal spring (top) and ideal dashpot (bottom).

In Figure 3-16 the step response of the strain x for a fixed force F is shown for an ideal spring (top) and dashpot (bottom). The spring reacts instantaneously on the force F with strain x equal to F_s/k_s . When the force is removed the strain decreases to zero. For a dashpot the behavior is different. Directly after the time t_0 when the force is applied the velocity v is equal F_d/k_d , the strain x increases linearly with time. At t_0 the strain x is zero, between t_1 and t_0 the strain is equal to the product $(v \cdot (t - t_0))$, when the force is removed the strain x is equal to $v \cdot (t_1 - t_0)$ and will remain at that value as long as no force is applied.

The first order differential equations of motion for a viscoelastic material with and without mass are solved in appendix XXX. When there is a mass oscillations and resonance phenomena occur, this is typical for realistic materials and structures. The last questions refer to structures with mass.

3.4 References

1. Biomechanics, mechanical properties of living tissues, Y.C. Fung, Springer, ISBN-13: 978-0387979472.
2. Medische fysica, A. van Oosterom and T. Oostendorp, Publisher: Bohn Stafleu van Loghum; 4th ed. 2016 edition, **ISBN-13:** 978-9036810852.

3.5 Questions

1. What is the difference between force and pressure?
2. Describe the three different types of pressure.
3. The SI unit of pressure is Pa. Calculate the conversion factors between mmHg, cmH₂O to Pa.
4. What is the difference between elastic and plastic deformation.
5. Two springs with spring constants k₁ and k₂ can be placed in series or in parallel combinations. Derive the effective spring constant of a replacement spring for the series and parallel combinations in terms of k₁ and k₂.
6. What is the Poisson ratio?
7. Derive the Laplace equation for a sphere with wall thickness d and radius R.
8. Two balloons with identical wall material are inflated to radius R₁ and R₂. R₂ is larger than R₁. The inflated balloons are connected to a tube with a valve between the two balloons. Describe what happens when the valve is opened.
9. What is stress and strain? What is the Young modulus?
10. Describe the stress-strain relation of a biologic tissue.
11. What is the elastic compliance? Is it constant for a biologic tissue?
12. What is the difference between a viscoelastic and elastic material?
13. What is hysteresis?
14. Describe the Voigt and Maxwell models, which model includes creep?
15. What is resonance? Which material parameters influence resonance?
16. Why is the parameter ζ important for a spring-mass-damper system?
17. Describe the difference of displacements of an undamped mass-spring system for a step force and a harmonic force at the resonance frequency.

4 FLUID TRANSPORT

In the circulation and respiration systems gases and blood are transported through elastic tubes towards tissues. Furthermore in many sensors that are used in the clinical domain fluid transport determines its frequency response and transfer characteristics. To understand and model the flow of fluids in tubes a basic understanding of fluid dynamics is needed. In this lecture fluid flow in straight tubes will be discussed briefly. Fluid dynamics is a very complex branch of physics, for a more detailed discussion see [1]. A more detailed description including derivations of some of the equations can be found in the appendices.

4.1 FLUIDS

Fluid dynamics is a part of mechanics that describes the flow of fluids due to driving forces. Fluid properties differ from the solid-state viscoelastic objects we have discussed so far. The differences are illustrated using Figure 4-1.

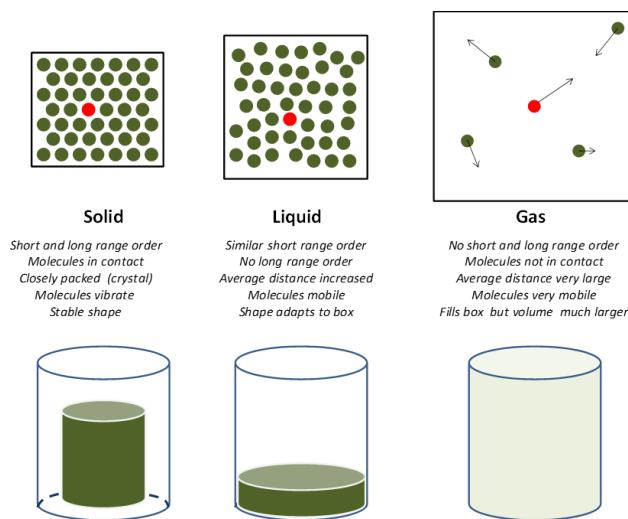


Figure 4-1 Schematic drawing of molecular position in solid crystal, fluid and gases phases including shapes of continuous macroscopic objects filled with a solid, a liquid and a gas.

Materials consist of a large number atoms or molecules, its properties depend on the attractive forces and temperature. In a crystalline solid-state material, atoms and molecules are on fixed positions. A block of solid material has a fixed shape and volume, the material does not flow. When an external force is applied to the object, the distance between the particles changes and a reaction force is generated. When the force is removed, the particles return to their original equilibrium position, the object returns to its original shape and size. This is elasticity and it was treated in chapter 3.

For temperatures above the melting point there is a phase change from solid to the liquid phase. In a liquid the atoms and molecules can move through the liquid. The higher the temperature the higher the particle mobility, the easier the liquid flows, the thinner the liquid. The density of the liquid is in general somewhat lower than the solid form, typically by only a few percent. Water is a rare exemption, it behaves different, below 4 °C the volume increases and density decreases. Therefore ice floats on water. The liquid volume is almost the same as that of the solid. During melting the liquid will flow and the object loses its shape. The liquid object shape adapts to the box in which it is contained. It has a free top surface that is in this case in contact with the ambient air.

When a force is exerted on a liquid, two cases can be considered. The first is the static case in which there is no net flow of material, for example when a liquid is compressed in a rigid container by a

piston. The other case is when net volume flow occurs, an example is a liquid in a tube when there is a net force applied to the liquid in the tube direction. When a net force is applied on the liquid via the tube openings, the liquid will start to move, the atoms or molecules will have a drift velocity component on top of the thermal random motion. Particles in the liquid are transported to another location in the tube and when the force is removed the transported particles stay there. Hence there has been a net transport of particles, there has been both volume and mass flow. The difference between volume and mass flow is important when liquids are compressible. When there is flow, the system is a dynamic system, this is the topic of fluid dynamics.

When the temperature is raised above the boiling point the thermal energy of the atoms or molecule will increase further, particles evaporate from the liquid and finally another phase transition occurs, i.e. the gas phase. The complete box in Figure 4-1 is now filled with the atoms/molecules. The density of the gas is several orders of magnitude lower than that of the liquid state. For atmospheric gas pressures the volume of the box should be more than 1000 times larger. The gas particles can move freely, but collisions with atoms/molecules and the walls of the container occur. The mean distance between particle collisions is a function of the gas pressure, note that this distance is much larger than the dimensions of the gas particles. The volume and mass flow in the gas phase can be described by similar equations as in the liquid phase. *Hence from the fluid dynamics point of view fluids can be both gases and liquids*, the main difference being the parameters (such as density and viscosity) used in the models.

4.2 FLUID PROPERTIES

The three most important parameters of fluids used in fluid dynamics are density, compressibility and viscosity (see Figure 4-2). Density and viscosity are the most important parameters for physiologic fluid transport.

The mass density ρ (or in short density) is defined as the mass per unit volume. The SI unit for density is kg/m^3 . This unit will be used in engineering calculations. This is not a practical unit in the medical domain, more commonly used units are: gr/cm^3 or gr/ml . Densities of liquids are orders of magnitude larger than those of gases. A list of values for relevant substances is given later in this section.

Compressibility is a relative measure of volume change of a solid, liquid or gas per unit applied pressure. For a solid or liquid, the compressibility is orders of magnitude smaller than that of gases. It is neglected for all solids and liquids in this course. For gases compressibility is much larger but since exerted pressures in the airway system are very small it can often be neglected for cases discussed in this course. When compression is important it will be mentioned specifically.

The viscosity is a very important parameter for fluid flow. The fluid viscosity causes flow resistance and subsequent dissipation of kinetic energy in the form of thermal energy that is lost and flows in the form of heat to the environment. Viscosity is defined as: "A quantity expressing the magnitude of internal friction in a fluid, as measured by the force per unit area resisting uniform flow". An alternative definition is: "The state of being thick, sticky, and semi-fluid in consistency, due to internal friction". Viscosity is caused by friction between adjacent fluid layers with different velocities. A large value of the viscosity causes the fluid to be thick and flow slowly as is illustrated in the right figure of Figure 4-2. Before discussing viscosity in more detail an important method of flow visualization is discussed.

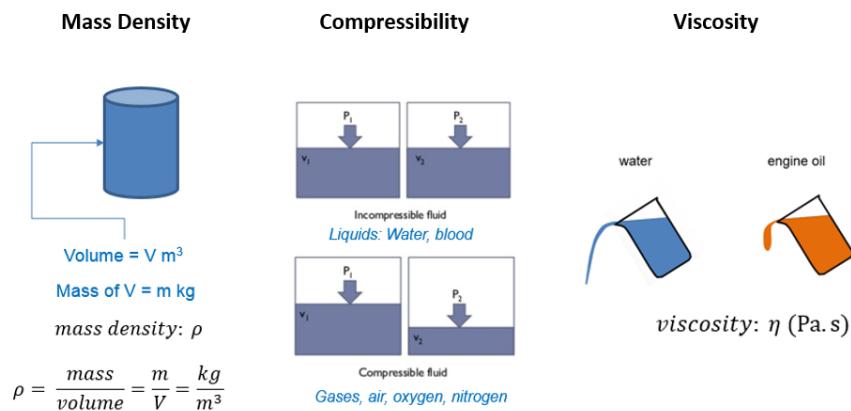


Figure 4-2 Three important fluid parameters, mass density, compressibility and viscosity.

4.2.1 Pathlines

A pathline (blue line) is the trajectory of an hypothetical small macroscopic fluid particle in the fluid (see Figure 4-3). At each point of the path line the particle has a velocity (red vectors) that is tangent to the pathline.

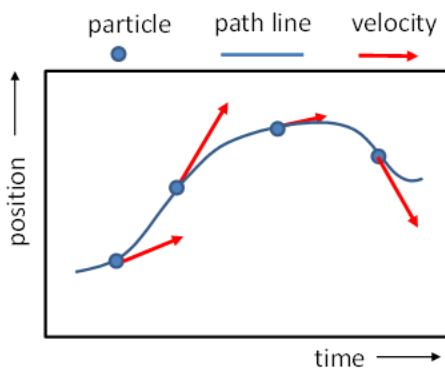
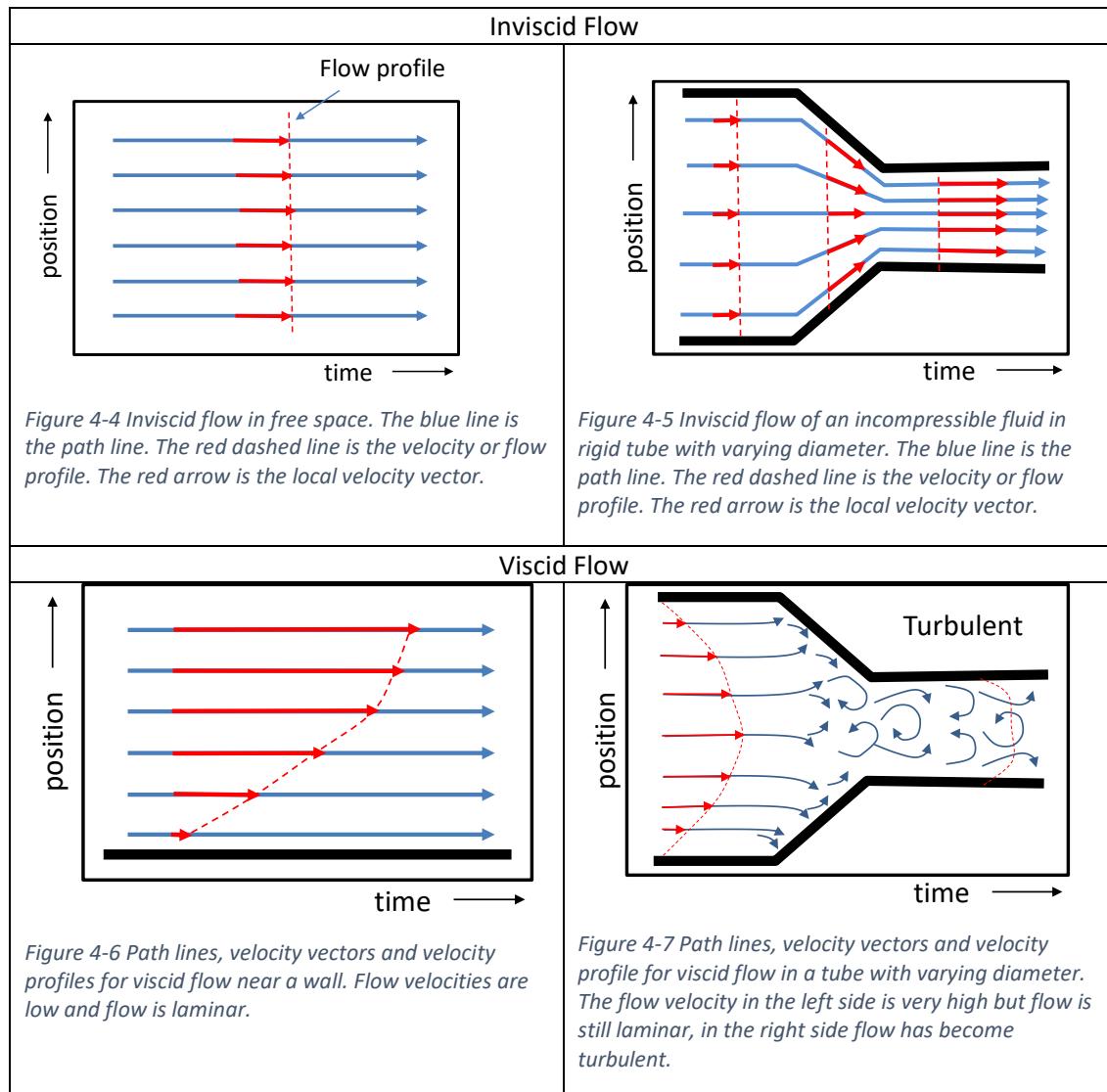


Figure 4-3 Schematic drawing of a path line and the position of a hypothetical fluid particle and the velocity vector at four instants of time.

In the figures below pathlines are shown for different types of flow for different geometries. In Figure 4-4 and Figure 4-5 path lines are shown for inviscid flow (no fluid viscosity or $\eta = 0$) for two different geometries. In free space the particles are moving in parallel trajectories and each particle has the same velocity. The flow profile is a plot of the velocity vectors at a fixed time over a cross sectional area that is perpendicular to the path lines. Hence the velocity profile is a plane. When the velocity profile is planar the flow is of the plug type (Figure 4-4). The particles move smoothly and follow regular paths that do not mix with each other. This type of flow is called laminar. In Figure 4-5 inviscid flow in a rigid tube with different tube diameters is shown. In the far left and far right the particle velocities are constant but in the larger diameter tube the particle velocity is lower than in the small diameter tube. In the transition area the velocity vectors are not parallel, particle velocity is increasing from left to right. This dependence on diameter is explained during the discussion of Bernoulli's law. The path lines are still smooth, regular and do not mix, hence the flow is of the laminar flow type.

In Figure 4-6 and Figure 4-7 path lines, velocity vectors and velocity profiles are shown for a steady viscous flow (i.e. viscous fluids with $\eta > 0$) for two geometries. In Figure 4-6 the path lines for viscous

flow near a wall are shown. The fluid velocity near the wall is zero due attraction and friction between the fixed wall and moving fluid particles. This is a so-called no-slip boundary condition. This no-slip boundary condition is used in the remainder of the course. The path lines and velocity vectors are parallel to the wall but due to viscous friction forces and the no-slip condition the magnitude of the velocity of particles increases when distance from the wall increases. The velocity profile is nonlinear, in this case it is parabolic.



In Figure 4-7 the path lines and velocity profiles are shown for steady flow at the inlet side for a tube with varying diameter in the lateral direction. At the inlet side fluid velocity is high but inflow is still laminar. In the wide tube the velocity profile is parabolic, velocity is zero at the wall and maximal in the center. When the tube diameter decreases fluid velocity increases but in the specific example the fluid velocity exceeds a critical value, the flow becomes turbulent. In turbulent flow instabilities in the flow pattern occur, eddies occur but the pattern changes continuously, the flow is chaotic and noisy, mixing occurs, energy dissipation increases. Turbulence may occur when the Reynolds number ($Re = \frac{\rho v D}{\eta}$) exceeds a critical value. This number is a ratio between inertial and viscous forces. The critical value Re is not a fixed number and depends on a lot of variables like geometry and wall smoothness. Note that turbulence occurs only in viscid flow and not in inviscid flow. Turbulence is

discussed in a separate section. In the following section the frictional forces, shear stress and shear rate that arise during flow in viscous liquids are discussed.

4.2.2 Viscosity Shear Stress and Shear Rate

In Figure 4-8 an experiment is described in which important terms and parameters are defined. The system consists of two large plates with area A and a separation d and a fluid between the plates with viscosity η . The separation d is much smaller than the plate dimensions.

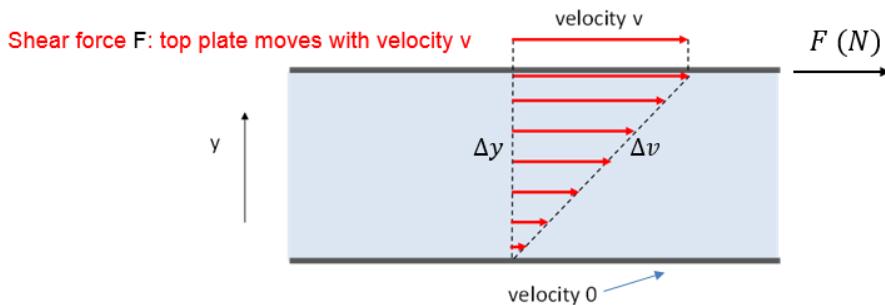


Figure 4-8 Steady state fluid velocity in the y-direction for a top plate moving with constant velocity v and a fixed bottom plate. The pathlines and velocity profile are shown in the figure. The force F is the shear force that is applied to the top plate.

A force F is applied to the top plate with area A in a direction along the surface plane of the fluid. This force F is called a shear force. The force F per unit area A ($\tau = F/A$) is the shear stress. The experiment shows that the shear force F is proportional to the top plate velocity v and plate area A and inversely proportional to the separation between the two plates d (equation 4.2-1). The proportionality constant is the viscosity η (equation 4.2-2). By measuring the force, velocity and knowing the dimensions A and d the viscosity of the fluid can be determined.

	$F \propto A \frac{v}{d}$	4.2-1
--	---------------------------	-------

	$F = \eta A \frac{v}{d}$	4.2-2
--	--------------------------	-------

The ratio v/d is the shear rate. For this example, the flow profile is linear. The friction force F between two very thin layers of fluid with y -coordinates differing by Δy and velocity differing by Δv is in the limit of infinitesimal small Δy determined next. In analogy with equation 4.2-2 we obtain equation 4.2-3 for the shear stress and shear rate in the limit of Δy towards zero:

	$\tau = \frac{F}{A} = \eta \frac{dv}{dy}$	4.2-3
--	---	-------

Equation 4.2-3 is valid for arbitrary flow profiles in tubes. The equation in cylindrical coordinates is equation 4.2-4:

	$\tau = \eta \frac{dv}{dr}$	4.2-4
--	-----------------------------	-------

The viscosity has dimensions Pa.s in the SI system, for the cgs system the unit is Poise, 1 Poise is equal to 0.1 Pa.s. In the above equations it is assumed that the viscosity is a constant that does not depend on shear rate, dimension and other factors. A fluid with constant viscosity, i.e. independent of dimensions and flow velocities, is a so-called Newtonian fluid. The viscosity and shear rate are very important physiological parameters.

4.2.3 Numeric values of viscosity and density

The numeric values of fluid density and viscosity of fluids commonly encountered in human physiology and in clinical sensing devices are shown in the table below.

Fluid	Density (kg/m^3)	Viscosity (mPa.s)
Water	1000	0.89
Blood plasma	1025	1.3
Blood (Hct = 40%)	1060	3
Motor oil	900	319
Glycerol	1259	1200
Air	1.2	0.0182
Oxygen	1.33	0.02
Nitrogen	1.2	0.018

The densities of the gases and liquids differ by almost three orders of magnitude. Differences in viscosity between liquids/gases are in the order of a factor of 50 or larger, hence liquids have much larger viscosity. When the viscosity is constant, independent from the shear rate and dimensions of the system the fluid is called a Newtonian fluid. In this course we will assume that fluids are Newtonian. Data from blood shown in Figure 4-9 show that blood is not a Newtonian fluid. This behavior will be ignored in this document. In the next section volumetric and mass flow rates are defined.

Physiological relevance

Warning. Water and blood plasma are Newtonian fluids, blood is not. Blood consists of blood plasma and particles in the form of red (disc like) and white blood cells. The red cells have a profound influence on viscosity which is illustrated in Figure 4-9. Viscosity of blood increases sharply with hematocrit of blood, decreases with shear rate (i.e. high viscosity at low flow rates) and decreases with diameter of the vessels down to 20 μm diameter. This is discussed in more detail in a lecture on human physiology.

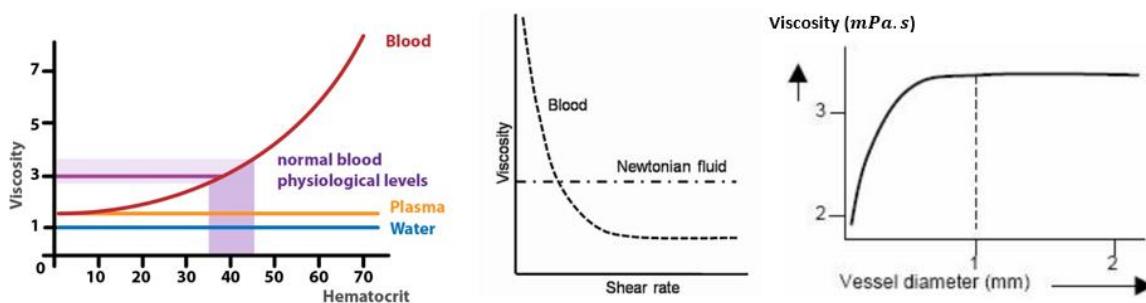


Figure 4-9 Viscosity of blood as function of the hematocrit (Hct: volume percentage of red blood cells), shear rate and blood vessel diameter.

4.3 Volumetric and Mass Flow Rates

It is assumed that flow is laminar, that the tube is rigid and that flow velocity profile is a plane. In Figure 4-10 volume flow rate is shown schematically.

The volume flow rate Q is defined as the volume that passes a cross section area A per unit time. The definition of Q is shown in Figure 4-10. It takes a time Δt before the volume V has passed the dashed cross section with area A . The dimensions of Q are m^3/s in SI units. In clinical use these units are not very practical and other units like ml/s or $\text{liter}/\text{minute}$ are used. The volume flow rate Q can also be

expressed in terms of area A and the flow velocity v_x as the volume is equal to the product of A and Δx the length of the cylinder of volume V and area A. This equation is simpler as flow velocity and area A can be measured with high accuracy. The equation that relates volume flow rate and flow velocity is valid only for a planar surface. In general Q is equal to the vector inner product ($\mathbf{v} \cdot \mathbf{A}$). For arbitrary surfaces equation 4.3-1 is used, Q is the surface integral of the velocity vector over the surface area A.

	$Q = \iint \mathbf{v} \cdot d\mathbf{A}$	4.3-1
--	--	-------

	$Q = \frac{\dot{m}}{\rho}$	4.3-2
--	----------------------------	-------

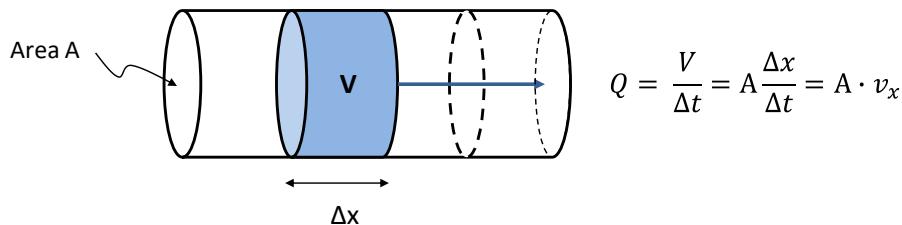


Figure 4-10 Schematic drawing of volume flow rate Q in a rigid tube with cross sectional area A.

Most liquids are nearly incompressible. This is also a good approximation for gases in the respiratory domain as pressures and flow velocities are very small and effects from compression can be neglected. In the remainder of the course it is assumed that fluids are incompressible unless mentioned otherwise. For an incompressible fluid the density ρ is constant and volume flow rate Q can be expressed in terms of the mass flow rate \dot{m} (kg/s) and density ρ . The mass flow rate is a better measure of flow for gases. For gases pressure and temperature may vary and volumetric flow will change. Mass flow is not dependent on these parameters. Many sensors are designed to measure the mass flow rate therefore equation 4.3-2 is important for practical use.

The definitions of flow and the types of flow are defined, the next step is to derive the basic equations of fluid flow in tubes.

4.4 FLUID DYNAMICS

Pressure-flow relations in tubes are derived from the Navier-Stokes equation (NS). This equation and the continuity equation are used to obtain equations for the relation between the driving forces and resulting volumetric flows. The NS equation is very complex and analytic solutions for many cases are either not possible or far too complex for the present course. A more detailed discussion and derivation of NS can be found in [1]. For practical reasons simple lumped element models are used in this course to model flow in tubes. These simple models provide insight in sensors, physiology and pathophysiology and are used in the majority of the available sensor and physiology models. In the following chapters flow-pressure relations for three basic lumped elements (resistor, inertance, compliance) are derived. The derivations use simplified forms of the NS equation. Before starting with the lumped element models a short summary of important conservation laws and laws of motion that are relevant for the models discussed followed by the NS equation for non-compressible flow in cylindrical tubes.

For tube flows described in this course several approximations are made in the Navier-Stokes equation, these are listed below:

1. Laminar flow only
2. Long straight pipes, edge effects can be neglected
3. Cylinder geometry
4. Incompressible fluid with density ρ
5. Newtonian fluid with viscosity η (viscid flow) or no viscosity (inviscid flow)
6. Gravitational effects are neglected

Note that these are crude approximations which are sometimes not valid for sensors and human physiologic systems. For pipe flow three limiting cases (basic lumped elements) are discussed, flow-pressure relations for idealized conditions are derived. Element 1: the resistor (R): flow where fluid friction is the dominant reactive force (neglect wall elasticity, fluid mass). Element 2: the compliance (C): flow into an elastic tube (neglect fluid mass and viscosity). Element 3: Inertance (L): flow velocity is determined by inertance of the mass of the fluid (viscosity, elasticity neglected). The driving force for flow is in all cases a pressure gradient in the axial direction x (see Figure 4-11).

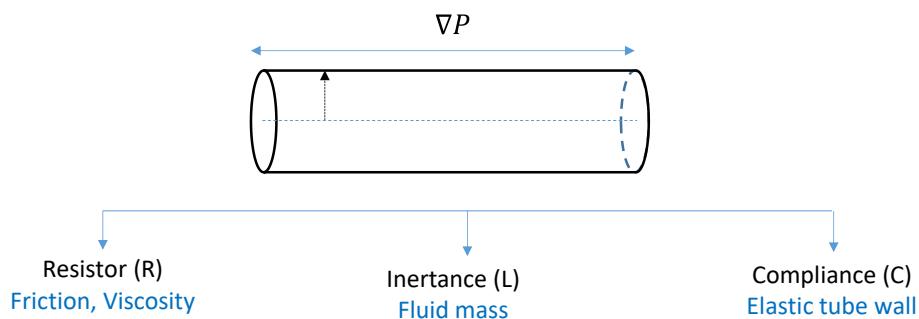


Figure 4-11 Three limiting cases of pipe flow, resistive flow, inertance dominated flow and flow into an elastic pipe. The driving force is proportional to the pressure gradient.

4.4.1 Viscid flow and Flow Resistance

The following additional approximations are made for derivation of the equation of the flow resistance:

1. Cylindric tube shape
2. Steady flow, no time dependence
3. Mass of fluid is neglected ($\rho=0$, no inertance)
4. Rigid wall material (i.e. no elastic deformation)
5. Radius r_0 and length l
6. No-slip boundary condition for flow velocity at the wall ($r=r_0$)
7. Boundary condition: maximum flow velocity at axial position ($r=0$)

For these approximations the Navier-Stokes equation (cylinder coordinates) reduces to:

	$\eta \left[\frac{\partial^2 u_x}{\partial r^2} + \frac{1}{r} \frac{\partial u_x}{\partial r} \right] = - \frac{\partial P}{\partial x}$	4.4-1
--	---	-------

Note that the velocity u_x in the x -direction is a function of the radius coordinate only, all other velocity components and gradients of these velocities are zero. The first step is to obtain the velocity profile $u_x(r)$. This is obtained by double integration of equation 4.4-1 and including the two boundary conditions 6 and 7. The result is:

$$u_x(r) = -\frac{\partial P}{\partial x} \frac{(r_0^2 - r^2)}{4\eta} = \frac{P_2 - P_1}{l} \cdot \frac{(r_0^2 - r^2)}{4\eta}$$

4.4-2

The velocity profile is parabolic and is shown in Figure 4-12. Integrating the velocity profile (equation 4.4-2) over the cross section (see Figure 4-12) gives an expression for the volumetric flow Q over the tube:

$$Q = \frac{\pi r_0^4}{8\eta l} \cdot \Delta P$$

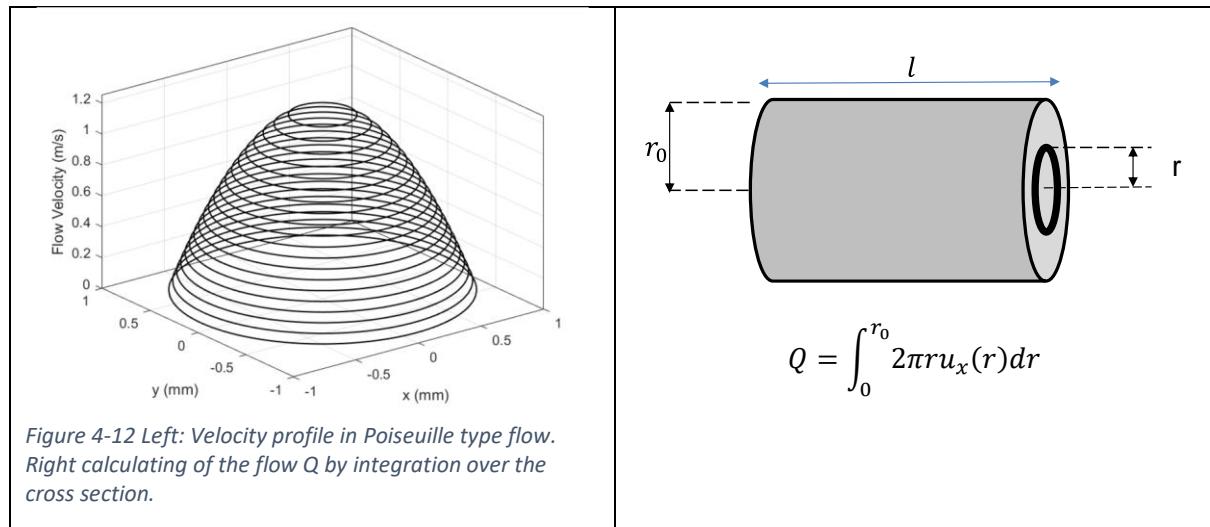
4.4-3

Equation 4.4-3 has the same form as Ohm's Law in the electrical domain, flow rate Q (i.e. current) is proportional to the pressure difference ΔP (voltage difference) and the proportionality constant is the inverse flow resistance R . In analogy with electrical resistance a flow resistance $R = \Delta P/Q$ is defined:

$$R = \frac{8\eta l}{\pi r_0^4}$$

4.4-4

This is the very important formula for flow resistance R , often called the Poiseuille resistance. Flow resistance is proportional to the viscosity and length of the tube and inversely proportional to the **fourth power** of the tube radius. This strong dependence on tube radius is due to the dependence on cross sectional area (same as in electrical domain) and the parabolic velocity profile (absent in the electrical domain for macroscopic wire dimensions). The dependence of resistance on radius is much stronger than the dependence of electrical resistance on wire radius. Note that this strong dependence on radius is very important for physiology and sensing systems, a decrease in radius with a factor of 2 results in an increase in resistance with a factor 16!



Physiological relevance

The dependence of tube resistance on radius is very strong and this is of enormous physiological importance for the regulation of blood flow to tissues. A small variation in radius leads to a large change in resistance. Furthermore, the dependence on viscosity should be noted as blood viscosity can vary, it depends on geometry as well.

It is apparent that the flow resistance is very small for large diameter tubes. Flow in such tubes the relation between flow rate and driving pressure gradient is dominated by other effects. For such large diameter tubes viscosity can then be neglected and the regime of inviscid flow is entered. In the

following two sections flow pressure relations are derived for tubes with inertance and elastic compliance dominated systems.

Summary

For a steady laminar viscous flow the volumetric flow Q is proportional to the pressure difference ΔP . The proportionality constant is the inverse flow resistance R . (Poiseuille)	$Q = \frac{\pi r_0^4}{8\eta l} \Delta P$
The flow resistance R is proportional to tube length l and viscosity η and inversely proportional to the fourth power of the tube radius r_0 .	$R = \frac{8\eta l}{\pi r_0^4}$
Flow Q and pressure difference ΔP are in phase with each other, flow changes instantaneously with changes in pressure gradient.	$\Delta\varphi = 0$
The fluid flow exerts a shear stress τ on the tube wall	$\tau = \frac{r_0 \Delta P}{2 l}$
Energy is dissipated in the form of heat (lost to the environment)	$E_{dis} = Q^2 R$

4.5 Inertance – Inviscid Flow

The following additional approximations are made:

1. Cylindrical tube
2. Time dependence included
3. Mass of fluid included (density ρ and volume of tube)
4. Inviscid flow, no friction in fluid and between fluid and wall
5. Newtonian fluid with viscosity equal to zero.
6. Rigid wall material (no elastic deformation)
7. Tube radius r_0 and length l

For these conditions and using the relation $Q = A \cdot u_x$ for flow the NS equation reduces to Newton's second law:

	$\rho \frac{\partial u_x}{\partial t} = \frac{\rho}{A} \cdot \frac{\partial Q}{\partial t} = \frac{\Delta P}{l}$	4.5-1
--	--	-------

This equation can be written as:

	$\Delta P = L \cdot \frac{\partial Q}{\partial t}, \quad L = \frac{\rho l}{\pi r_0^2}$	4.5-2
--	--	-------

Note that equation 4.5-2 that relates flow with a pressure gradient in the mechanical domain is similar as the current-voltage relation of an inductor in the electrical domain. The parameter L in equation 4.5-2 is called inertance and it is proportional to fluid density, length of the tube and inversely proportional to the cross sectional area A of the tube. The inertance in the fluid dynamic domain follows directly from the second law of Newton ($\Sigma F = m \cdot a$).

The pressure difference ΔP is proportional to the rate of change of the flow rate Q . For harmonic driving pressure there is a frequency dependence of the impedance of element and there is a phase shift between flow and pressure waveforms. This is illustrated in Figure 4-13. Given a pressure waveform the resulting flow can be calculated by integration of equation 4.5-2. Due to inertia (i.e. the mass of the fluid in the tube) the changes in flow (motion) follow changes in pressure (net driving force). *Hence flow can only change after a pressure change has occurred (change in driving force).*

Stated differently pressure changes lead flow changes. There is a positive phase difference between pressure and flow. In the left figure there is a pressure pulse applied over the tube, the resulting flow

is obtained by integration. Flow increases when the pressure pulse is positive and becomes constant for longer time when the pressure pulse is zero. A phase difference between pressure and flow is observed. Note the constant flow after larger time when the pressure pulse is zero. This follows from the first law of Newton. The net force on the fluid is zero and the fluid velocity and flow remain constant at its maximum value⁵. Note that kinetic energy is stored in the inertance L, there is no energy dissipation. The kinetic energy E stored in the moving liquid mass is equal to:

$$E = \frac{1}{2} L Q^2 \quad 4.5-3$$

In the second plot (right Figure 4-13) another characteristic feature in the flow waveform of an inertance element is shown. *Note that the flow is positive over the entire interval* (forward flow), initially flow increases ($\Delta P > 0$), reaches a maximum ($\Delta P = 0$) and decreases thereafter ($\Delta P < 0$). The pressure gradient is positive in the leading edge of the forward flow pulse and negative for the trailing edge. *Hence the pressure gradient reverses sign while there is still forward flow.* This behavior follows from the second law of Newton. Initially the fluid mass is accelerated, flow increases, this requires a positive net force (pressure gradient). In the trailing edge the flow decreases, the fluid velocity decreases, this is caused by a negative net force gradient. For an inertance element there can be (for a short while) fluid flow in the forward direction after the pressure gradient has switched sign from positive to negative.

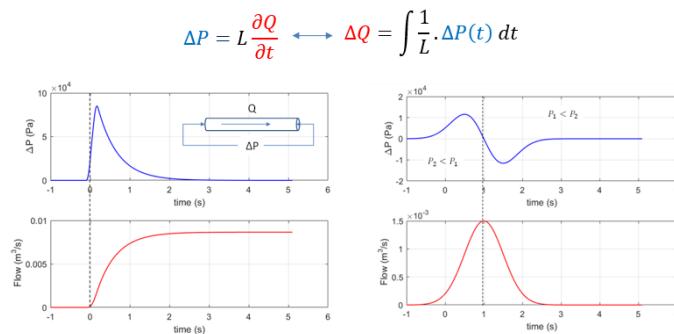


Figure 4-13 Pressure and flow plots versus time for two different pressure waveforms. The calculation of flow is shown above the figure.

Summary

<i>Inertance L is related to the acceleration or deceleration of the fluid mass in a tube, it follows directly from Newton's second law. It depends on fluid density and tube dimensions.</i>	$L = \frac{\rho l}{A} = \frac{\rho l}{\pi r^2}$
<i>Inertance L scales with $1/r^2$ (r is the radius of the tube)</i>	
<i>In an inertance element flow Q changes after a pressure gradient is applied. In an inertance element flow can only change with finite speed.</i>	
<i>For an harmonic pressure wave pressure and flow are not in phase, ΔP change (the cause) leads change in flow Q (the effect).</i>	$\Delta\phi = 90^\circ$

⁵ Note that this behavior is not observed in practice because fluids have also flow resistance.

<p><i>Pressure and flows are related by the equation in the right column. There is an analogy with an inductor in the electrical domain.</i></p>	$\Delta P = L \frac{\partial Q}{\partial t}$
<p><i>In an inertance element energy is stored in the form of the kinetic energy of the fluid. In an inertance kinetic energy is conserved.</i></p>	$E_{kin} = \frac{1}{2} L Q^2$

Physiological relevance

In the last phase of blood ejection from the left ventricle the pressure in the downstream area is higher than that in the ventricle. This is caused by the inertance of blood, blood velocity decreases but there is still forward flow. A second example is that of pressure- and flow-waves in the arterial tree that is treated in section 4.7.

In the following section the pressure-flow relation for a tube with an elastic wall (compliance) is discussed.

4.6 Flow in an elastic tube – Compliance and Tube Collapse

The following additional approximations are made:

1. The fluid has no mass (non-inertial flow)
2. The fluid has no friction (inviscid flow)
3. The wall material consists of an ideal elastic material
4. The compliance can depend on the transmural pressure
5. The pressure difference along the tube is much smaller than the transmural pressure
6. Time dependence is included

The pressure-flow relations *cannot* be derived from the Navier-Stokes equation. A separate equation is needed, a simplified equation is derived below. A short recapitulation of volume-pressure relation of an elastic tube⁶ is shown in Figure 4-14. At time t_1 the volume is $V(t_1)$ and the internal pressure is $P_i(t_1)$. The tube volume is a function of the transmural pressure $P_{TM}(P_i - P_{ext})$ and it may depend on physical dimensions and elastic properties of the tube wall. Between times t_1 and t_2 there is a net inflow of fluid and the volume expands to $V(t_2)$, the internal pressure increases to $P_i(t_2)$ and transmural pressure increases.

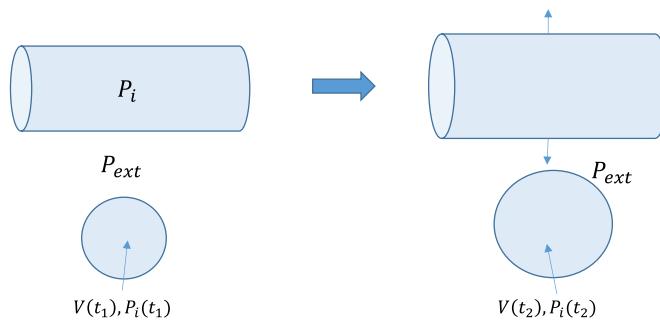


Figure 4-14 Elastic tubes after a volume change between times t_1 and t_2 for a fixed external pressure.

The volumetric flow-pressure relation is derived using the definition of volumetric flow (equation 4.3-1). The tube volume is proportional to the magnitude of the transmural pressure, the time

⁶ See appendix 12.1 on pressure-volume relation of elastic tubes

dependence of the volume is related to changes in time of the transmural pressure. This leads to equation 4.6-1. Note that C_d is the differential elastic compliance at pressure P_{TM} i.e. $\left(\frac{dV}{dP_{TM}}\right)_{P_{TM}}$ of the tube and can depend on transmural pressure. The compliance does not depend on time.

$$Q = \frac{\partial V(P_{TM})}{\partial t} = \frac{\partial V(P_{TM})}{\partial P_{TM}} \cdot \frac{dP_{TM}}{dt} = C_d \frac{dP_{TM}}{dt} \quad 4.6-1$$

The relation between flow Q and transmural pressure is the same as the current-voltage relation of a capacitor in the electrical domain. Two examples of pressure and volume versus time are shown in Figure 4-15 for a step change in the net inward flow for a harmonic flow wave. The pressure curves are obtained by integration of the flow. The change in transmural pressure always lags the change in flow (i.e. volume change) and transmural pressure can change only after the flow has started.

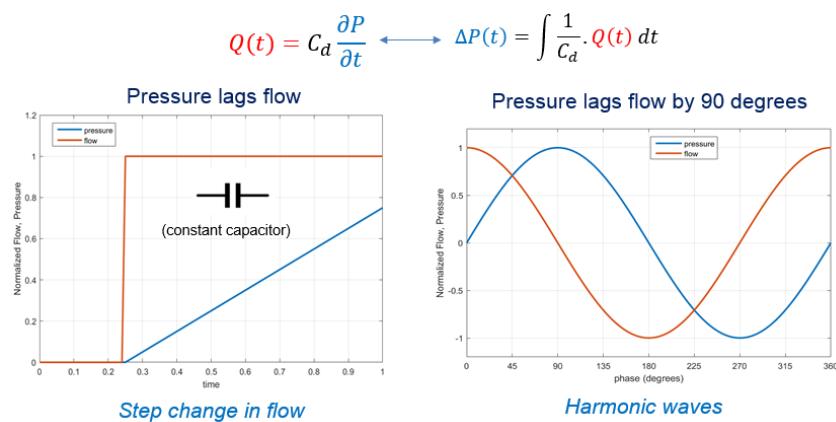


Figure 4-15 Pressure and volume versus time plots for a step and harmonic flow input.

When the volume of the elastic tube changes energy is stored in the form of elastic potential energy (i.e. similar to a spring), the potential energy E stored is equal to:

$$E = \frac{1}{2} C P_{TM}^2 \quad 4.6-2$$

Physiological relevance

The elasticity of the aortic wall limits changes in blood pressure during and after injection of blood from the heart. The elastic energy stored in the compliance is released during diastole and is converted into kinetic energy of blood. The elastic properties of the aorta are very important. The workload of the heart is reduced, and peak blood flow is reduced. Furthermore, the elastic compliance one of the factors responsible for the wave-like pressure and flows in the arterial tree. Note that this is a very efficient manner of energy transport.

Summary

<i>Compliance C is related to the wall elasticity of a tube, it follows from the Laplace and Hooke relations. It depends on tube dimensions, on wall Young Modulus and wall thickness.</i>	$C = \frac{dV}{dP} = \frac{2\pi r_0^3}{Ed}$
<i>Compliance of a thin walled tube scales with r_0^2 and r_0/d</i>	

<i>In a compliance element flow Q leads changes in pressure gradient. In a compliance element pressure can only change with finite speed.</i>	
<i>Pressure and flow harmonic waves are not in phase. Change in flow Q (the cause) leads change in ΔP (the effect).</i>	$\Delta\varphi = 90^\circ$ (harmonic waves)
<i>Pressure and flows are related by the equation in the right column. There is an analogy with a capacitor in the electrical domain.</i>	$Q = C \frac{\partial P}{\partial t}$
<i>In a compliance element energy is stored in the form of the elastic energy of the wall. In a compliance energy is conserved.</i>	$E_{el} = \frac{1}{2} CP^2$

4.6.1 Viscous flow through an elastic collapsible tube

In a rigid tube the flow rate is determined by the pressure difference between the inlet and outlet. The viscous flow through an elastic tube is more complicated, three pressures are of importance, the inlet and outlet pressures and the external pressure. The flow is determined by all three pressures via the driving pressures and the transmural pressure that varies over the tube length (determines the diameter over the length of the tube). *When there is flow through the tube the transmural pressure varies over the tube length.* Hence the local tube diameter and local resistance vary over the tube length. The strong dependence of resistance on tube diameter causes a non-linear pressure distribution over the tube length. The qualitative flow-pressure(s) relation of an elastic tube is discussed below. The following additional assumptions are made.

1. The fluid has no mass (no inertance)
2. Only steady state conditions are considered (static)
3. The fluid is Newtonian, friction and viscosity are included
4. The tube wall is an ideal elastic material with no mass
5. The tube cross section remains circular, depends on transmural pressure and is very small (but non-zero) for zero transmural pressure

A schematic drawing of the tube with flow Q at given pressures is shown in Figure 4-16. For this case the external pressure is smaller than both input and output pressures in the tube as indicated in the figure. The diameter of the tube decreases towards the output side. This is caused by a lower internal pressure near the outlet. Therefore both the local transmural pressure and tube diameter are smaller near the outlet side. The fluid resistance causes a gradual and strong non-linear pressure drop from inlet to outlet. The transmural pressure, tube cross section and local tube resistance become functions of the x-coordinate, the diameter decreases from input to output, the local resistance increases strongly with decreasing diameter. Note that tube resistance scales with the inverse of the fourth power of the tube radius. Hence the local tube resistance and pressure drop is non-linear with axial coordinate x and is highest near the outlet side. In summary due to the elastic tube wall, the tube diameter decreases from the inlet to the outlet side where it is smallest. For a given external pressure the tube resistance depends on both inlet and outlet pressure and is non-linear with respect to the pressure difference between inlet and outlet.

$$P_{ext} < P_{out} < P_{in}$$

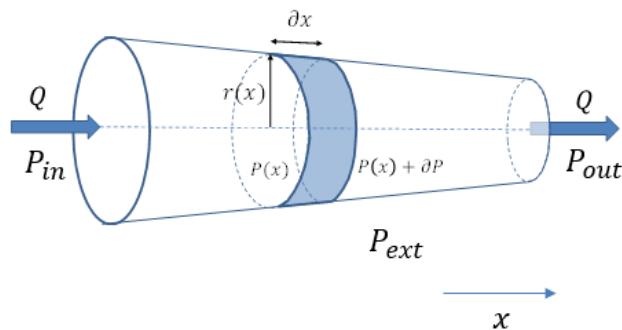


Figure 4-16 Elastic tube with flow Q for external pressure smaller than both input and output pressures.

A qualitative discussion of different flow regimes of an elastic tube for different pressure regimes is given below. This is illustrated in Figure 4-17 where four cases are shown with the specific flow included in the flow-pressure characteristics plotted below the tube geometries. The inlet pressure and external pressure are constant and are the same in the first three figures, the outlet pressure is increased in steps until it is smaller than the external pressure.

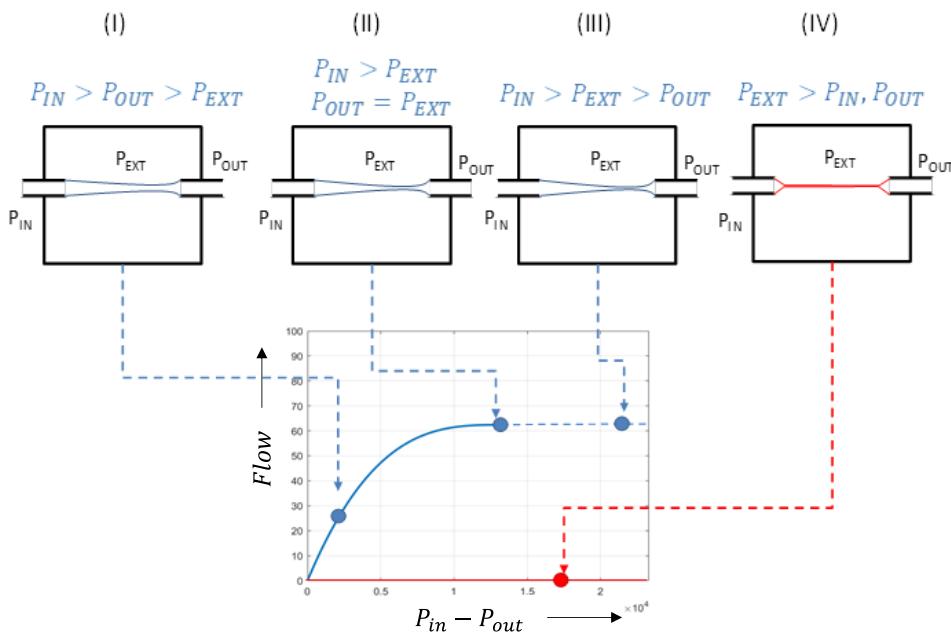


Figure 4-17 Four conditions of tube geometry with variable external pressure and constant inlet and external pressure (left three figures) indicated in the figure. The corresponding flows are indicated the flow-pressure characteristics below. Condition IV is a separate case with external pressure larger than tube pressures.

Condition I: The resistive regime. Tube pressures are much larger than external pressure. Flow increases with increasing pressure difference over the tube (decreasing outlet pressure). The increase in flow is not linear due to the increasing tube resistance near the outlet side.

Condition II: Maximum flow. The outlet pressure is further reduced until the outlet tube pressure approaches the external pressure, due to the large pressure drop at the outlet the transmural pressure approaches zero and the local tube cross section is very small (hence very high resistance). The tube collapses at the outlet side (i.e. reaches a very small but finite radius) and tube flow has

reached its maximum value. **The tube pressure at the outlet when collapse occurs is almost equal to the external pressure.** The pressure difference between inlet side and tube near the outlet has reached its maximum value.

Condition III: Flow saturation or waterfall regime. Further reduction of the outlet pressure does not lead to an increase in flow, i.e. flow is constant. The tube pressure near the collapse point of the tube remains roughly equal to the external pressure and this does not change with the reduction of outlet pressure as long as the inlet pressure is larger than the external pressure. The excess pressure is distributed over the tube region beyond the collapse point. The flow is determined by the difference between inlet and external pressure (which is constant), this difference does not change anymore with decrease in outlet pressure. There is an analogy with a waterfall. The flow is determined by the supply of water from the river, the height of the waterfall has no influence on the flow.

Condition IV: Collapse. When external pressure is larger than both inlet and outlet pressure the tube is collapsed over its entire length. Hence tube diameter is very small and the tube resistance is extremely large and flow (red curve) is negligible small.

Collapse and flow saturation occur in the circulatory and respiratory systems and will be discussed in more detail in other lectures.

Physiological relevance

Collapse of elastic tubes is of great importance when transmural pressure becomes negative in a physiological system. Examples are collapse of airways during forceful exhalation, in obstructive lung diseases, collapse of blood vessels in the pulmonary circulation and also in measurements such as the cuff based oscillatory blood pressure. When the flow in an elastic tube does not depend on the outlet pressure tube collapse is probably of importance.

4.7 Inviscid flow in long elastic tubes – transmission line

In the previous sections tube flow has been modeled for either resistive, inertance or elastic tubes that can be modeled by a single lumped element. In long larger diameter elastic tubes, flow resistance is finite but very small, walls are elastic (i.e. compliance) and tube fluid has mass (i.e. inertance). This combined effect of compliance and inertance leads to a wave propagation behavior that is very important for the physiology of the circulation system and in sensing systems. In this section the time dependent pressure and flow waves in long, large-diameter fluid filled elastic tubes are discussed. For long and larger diameter elastic tubes both inertance L and compliance C are dominant and flow- and pressure waves along the tube are observed. Fluid transport in this case differs appreciably from the simple lumped element models discussed before and continuum models are required.

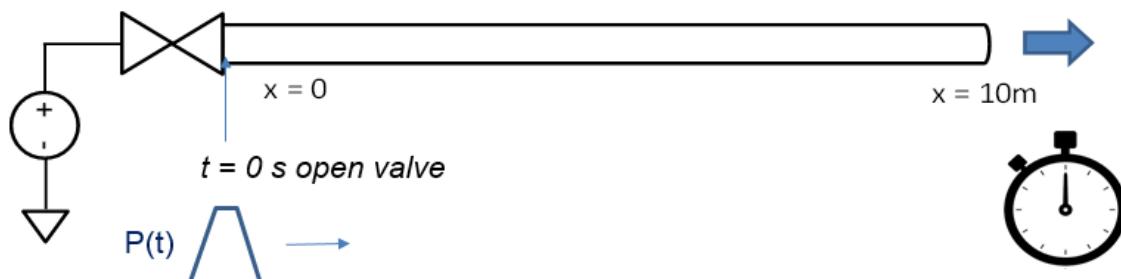


Figure 4-18 Schematic drawing of long tube experiment.

A simple experimental set-up shown in Figure 4-18 is used to illustrate flow and pressure induced by a short pressure pulse applied at the inlet. A long tube with length L of 10 meter and with radius of 1 cm is filled with water and is connected to a pressure source via an electronic valve that can be opened and closed in very short time. At the outlet of the tube an electronic pressure sensor is connected. We want to study the output pressure and flow pulses as function of time after the valve is opened. The trigger pulse of the valve and sensor output signal allow very accurate timing accuracy (ns range). Two tube materials are tested, one rigid stainless-steel tube and one elastic rubber tube.

The measurements on both tubes show that directly after the valve is opened there is no signal is observed at the output sensor. For the steel tube a pressure pulse is detected at the outlet after a delay of 20 ms, for the rubber tube the pressure pulse is detected after 2 seconds. Before the detection time there is no sign for the observer at the outlet that the valve at the inlet was opened. Note that the models developed for the resistor, inertance and compliance elements predict changes of the output pressure pulse directly after the valve is opened. These models of the basic lumped elements are not in agreement with the measurement results observed for the long tube as discussed above.

It is known that in long elastic tubes the wall elasticity in combination with the inertia of the fluid lead to pressure and flow waves. These phenomena are very important for the physiology of the circulation, for new blood pressure measurement techniques and for some sensors. In fluid dynamics a wave is an oscillation that transfers energy with a propagation velocity through a medium from one point to another point in space. Particles in the transmission medium are displaced rapidly from their position and collide with neighbors which again collide with their neighbors and so on. In an electrical transmission line the voltage pulse can travel with almost the speed of light, the individual electrons have velocities that are a factor 3000 smaller. Hence the propagation velocity of the energy pulse is much larger than that of the individual particles. The model for the wave propagation is a continuum model in the mechanical domain, it is analogous to that of a transmission line in the electrical domain. The derivation of a model for a lossless transmission line and one with losses is described in more detail in appendix 13.3.

The main features of the model for wave propagation are described using Figure 4-19. Since the tube diameter is large the tube resistance can be neglected (ideal tube). The wave behavior depends on the inertance and compliance of the tube. The first observation from the experiment is that the tube behaves as a single element with a real characteristic impedance Z_0 , the characteristic impedance does not depend on the tube length or frequency of the pulse. The relation between pressure and flow (equation 4.7-1) is identical to Ohm's law and since Z_0 is a real number there is no phase difference between pressure and flow waves:

	$Q = \frac{P}{Z_0}$	4.7-1
--	---------------------	-------

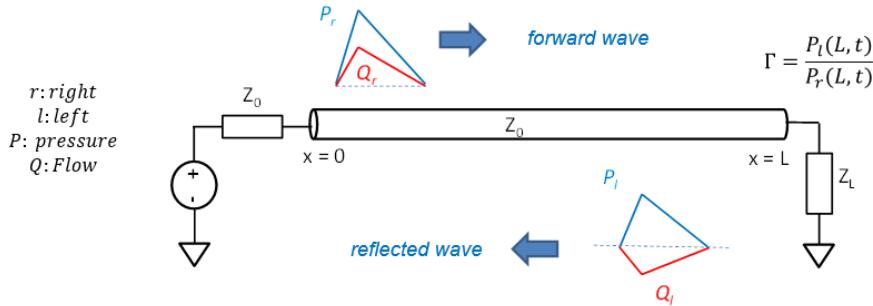
Hence the long tube impedance is like that of an resistor, however in contrast to a resistor there is no energy dissipation in the ideal tube. Since the fluid resistance is neglected, the waves are not attenuated⁷. This is an important observation; ideal transmission lines can transport energy and fluids without dissipation losses, i.e. they are ideal when energy needs to be transported over large distances with low energy losses. This is very important both in the engineering and physiology fields. The characteristic impedance for the ideal tube is given by:

⁷ In real tubes there is some resistance in the fluid and there are visco-elastic losses in the wall material. In this case damping occurs

$$Z_0 = \sqrt{\frac{L}{C}}$$

4.7-2

In equation 4.7-2 L and C are the inertance and compliance per unit length.



The pressure and flow waves are moving towards the right (i.e. outlet side) with a wave propagation speed⁸ v_p . This wave propagation speed does not depend on frequency. The pulse shape does not change during its travel through the tube. The wave propagation speed is given by:

$$v_p = \frac{1}{\sqrt{LC}}$$

4.7-3

The wave propagation velocity v_p is also called the pulse wave velocity (PWV). A first order equation for the pulse wave velocity can be derived by using expressions for tube inertance and tube compliance derived before. This gives the following expression for PWV.

$$PWV = \sqrt{\frac{E \cdot d}{2\rho r_0}}$$

4.7-4

Here E is the incremental Young modulus, d is the wall thickness, ρ the fluid density and r_0 the radius of the tube. After a time $t_L = L/v_p$ the first wave has reached the tube outlet which is now connected to a load with impedance Z_L . When the load impedance differs from the characteristic impedance of the tube the wave traveling into the right direction is reflected at the end of the tube. The reflected pressure and flow waves will travel to the left direction towards the pressure source. The reflected pressure wave has the same sign as the forward wave. The reflected flow wave has the opposite sign (direction of velocity is in the opposite direction). The reflected waves can interfere with forward waves on its trip to the source. The reflection coefficient Γ is defined as:

$$\Gamma = \frac{P_r(L,t)}{P_l(L,t)} = \frac{Z_L - Z_0}{Z_L + Z_0}$$

4.7-5

When the load impedance is equal to the tube characteristic impedance the reflection coefficient Γ is equal to zero, and no reflection occurs. For other load impedances reflection occurs and the

⁸ This is the same as the phase velocity in this ideal case.

pressure and flow waves in the tube are affected by interference between the forward and reflected pressure and flow waves. When reflection occurs the impedance of the tube is not equal to the characteristic impedance anymore. The impedance Z is defined by the local ratio of P/Q and this value will depend on position in the tube and length of the tube.

Using equation 4.7-4 the experimental results described earlier can be understood. The Young modulus of the rubber tube is much smaller than that of a steel tube and therefore a PWV is smaller. For the rubber tube the estimated PWV is around 5 m/s and a travel time to the end of the tube around 2 seconds is expected. The Young Modulus of steel is about ten thousand times larger than for rubber. For the steel tube PWV is 100 times larger (500 m/s) and the travel time is around 20 ms. This agrees well with the observations. No signal should be detected before time t_L the travel time of the pulse to reach the outlet. For most materials the Young modulus is a function of the transmural pressure. Measurement of PWV would then be a surrogate measurement of the transmural pressure. In appendix 13.3 transmission line models are discussed in more detail.

Physiological relevance

The arterial tree is a complex system of tubes with transmission line wave propagation of pressure and flow pulses. This is a very efficient transport mechanism. Reflections at bifurcations and near the periphery lead to increased systolic blood pressure (age related hypertension) and strongly distort the arterial pressure and flow waveforms.

In the previous sections it was assumed that tubes have cylindrical shape and gravity was ignored. However, the impact of gravity on flow and pressures is considerable, in the following section this is discussed in more detail and Bernoulli's equation is derived. This equation is very relevant for many clinical measurements and deserves careful consideration.

4.8 Bernoulli's Equation

The Bernoulli equation is very important for both physiology and sensors, it is valid only for inviscid stationary flow (large diameter tubes) and is derived for the following conditions:

1. Stationary flow
2. Incompressible flow
3. Inviscid flow
4. Laminar flow
5. Rigid wall material

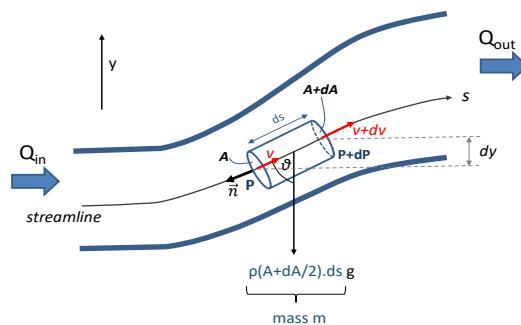


Figure 4-20 Tube with variable cross section and with inclination in the y-direction. A streamline and a control volume element with corresponding parameters are included in the figure.

In Figure 4-20 a schematic drawing of a tube with flow Q is shown, both gravitational effects and changes in tube cross section are important. A streamline s is shown. Euler's equation is solved along

the streamline s . The velocity vector is always tangent to the streamline, this simplifies Euler's equation to:

$$\rho \frac{\partial v}{\partial t} + \rho v \frac{\partial v}{\partial s} = - \frac{\partial P}{\partial s} - \rho g \cos \vartheta \quad 4.8-1$$

Flow is stationary and for the control volume at coordinate s the cosine term can be approximated by $\delta y/\delta s$, this gives the left term which can be rewritten as the right term:

$$\rho v \frac{\partial v}{\partial s} = - \frac{\partial P}{\partial s} - \rho g \frac{\partial y}{\partial s} \rightarrow \frac{\partial}{\partial s} \left(\frac{1}{2} \rho v^2 + P + \rho gy \right) = 0 \quad 4.8-2$$

Integration of the right term gives the Bernoulli equation 4.8-3:

$$\frac{1}{2} \rho v^2 + P + \rho gy = \text{constant} \quad 4.8-3$$

This equation is valid for all streamlines in the tube. P is the internal pressure due to collisions of the molecules with the tube wall, $\frac{1}{2} \rho v^2$ is the dynamic pressure term that is related to the collective flow velocity component v of the fluid and the term ρgy is the hydrostatic pressure component caused by the weight of the fluid at coordinate y . The three terms have the dimension of energy per volume or energy density. An interpretation of this very important equation is that it represents the conservation of energy density. The sum of the three terms is a specific constant (i.e. total pressure) which has the same value throughout the system:

$$\frac{1}{2} \rho v_1^2 + P_1 + \rho gy_1 = \frac{1}{2} \rho v_2^2 + P_2 + \rho gy_2 = \text{constant} = P_{\text{total}} \quad 4.8-4$$

Bernoulli's equation is an extremely important equation and it is used in many branches of physics, sensing, physiology and pathophysiology. This will come back many times in following chapters. Two examples of the great practical importance of this equation are discussed below. Note that the internal pressure is the pressure of interest, whenever the hydrostatic or dynamic pressures change the internal pressure changes as well (but in the opposite direction).

In the first example a tube of uniform cross section filled with water is considered (see Figure 4-21). The inlet and outlet are different heights, the measured inlet pressure P_1 is 10000 Pa. the flow velocity is 1 m/s. We would like to know the internal pressure P_2 at the outlet position. Since volumetric flow and tube cross section are the same in the tube the flow velocity is the same at inlet and outlet. Equation 4.4-4 then reduces to:

$$P_1 - P_2 = \rho g(y_2 - y_1) \quad 4.8-5$$

The hydrostatic pressure term on the right ($\rho g(y_2 - y_1)$) is equal to +9800 Pa, hence the pressure P_2 is +200 Pa. Note that a pressure sensor measures the internal pressure. The hydrostatic pressure effect of only one-meter height difference is of comparable magnitude as the internal pressure at the

inlet and cannot be ignored. This will come back in the physiology of the circulation where similar pressures and height differences occur.

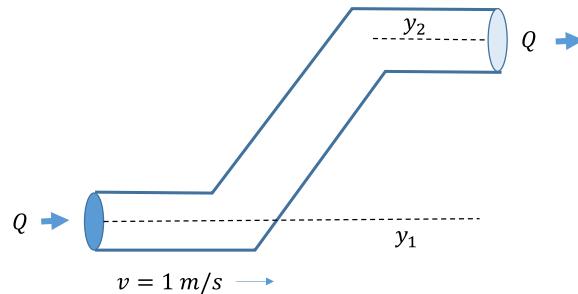


Figure 4-21 Water filled tube with uniform cross section with height difference y_2-y_1 of 1 meter between inlet and outlet. The flow velocity is 0.01 m/s. The gauge pressure at the inlet is 10000 Pa.

A second somewhat more complicated example is a nozzle (see Figure 4-22). It is a conical tube in which a constant flow Q is forced from inlet to outlet. It is used to accelerate fluids. We would like to know the difference in flow velocity and internal pressure between the inlet and outlet.

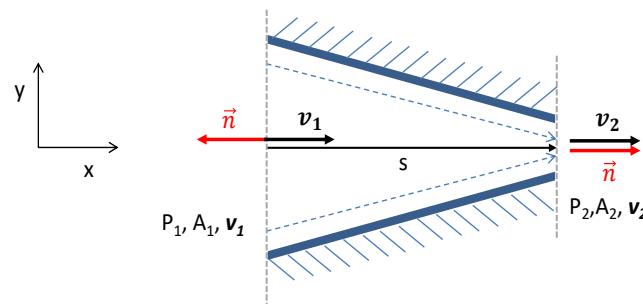


Figure 4-22 Nozzle with inlet and outlet cross section, flow velocity and internal pressures indicated in the figure.

This problem is slightly more complicated as the previous one since two steps need to be made. The first step is the estimation of the flow velocities. We know that the flow Q is the same throughout the tube, the fluid is incompressible, and this is a steady state flow. This can be solved by using the continuity equation for mass density. In a simplified form this equation can be stated as flow Q into the system must be equal to flow out of the system. This gives:

$$Q_{in} = Q_{out} \rightarrow A_1 v_1 = A_2 v_2 \rightarrow v_2 = \frac{A_1}{A_2} v_1 \quad 4.8-6$$

Since the flow and cross sections can be measured with high accuracy the inflow and outflow velocity can be measured. The outflow velocity is larger than the inflow velocity it is determined by the ratio of the inlet and outlet cross section areas. So, if the design aim is to achieve a certain outflow velocity, we know how to do it. The next task is to estimate the difference in internal pressure between inlet and outlet, the inlet pressure should be sufficiently large to maintain flow. This can be estimated with the Bernoulli equation. As mentioned before, we can use any streamline for this estimation, but the smartest streamline is the axial streamline as gravitational effects are absent.

This gives:

	$\frac{1}{2}\rho v_1^2 + P_1 = \frac{1}{2}\rho v_2^2 + P_2 \rightarrow P_1 = P_2 + \frac{1}{2}\rho(v_2^2 - v_1^2)$	4.8-7
--	--	-------

Equation 4.8-7 shows that the internal pressure at the inlet is larger than that at the outlet despite the larger tube cross section. This is called the hydrodynamic paradox and is consequence of the energy conservation law. Since the total pressure in the Nozzle is constant and gravity effects are absent an increase in dynamic pressure must be balanced by a decrease in internal pressure, i.e. the kinetic energy of the molecular motion is converted in part to collective flow motion. A mechanical engineer wants to design a nozzle such that the outlet airflow velocity is 100 m/s, the inlet flow velocity is 1 m/s. This is easy, the cross section at the outlet should be 100 times smaller than that at the inlet. The next question is: what should be the inlet pressure if the outlet pressure is atmospheric pressure? This is also straightforward, using equation 4.8-7 we find that the inlet pressure should be 6000 Pa higher than atmospheric pressure. This is 6% of atmospheric pressure and it is an acceptable pressure (low cost compressor is possible).

The above examples illustrate the practical importance of the Bernoulli equation, as mentioned before it appears in many engineering fields, for instance the lift of a wing of an airplane is for a large part generated by difference in flow velocity above and below the wing. However before using this equation it must be clear that fluid viscosity plays a minor role, when this is not true the equation may not be used, it is not valid anymore.

In the previous example a Nozzle was designed to generate output flows of 100 m/s. These are very high flow velocities and the assumption of laminar flow may not be valid anymore. The flow regime can change to the turbulent flow which is discussed in the next section.

Physiological relevance

Gravity has a large impact on blood pressure as well as the accuracy of blood pressure measurements. Furthermore, in the pulmonary circulation it causes non-homogeneous ventilation and circulation. Other examples from the circulation are flows near heart valves and stenosis in large blood vessels.

4.9 Turbulence

In previous sections laminar flow was discussed for flows where either resistive or inertance components are present. These models are valid for small flow velocities. When flow velocity increases the balance between inertial forces and viscous forces is lost. The inertial forces increase rapidly, and viscous forces cannot damp motions of fluid particles in the radial and azimuthal directions. Irregularities in fluid flow start to occur when flow velocity increases, with a further increase in flow velocity the flow becomes turbulent (see Figure 4-23 , Figure 4-24, Figure 4-25). Turbulent flow is characterized by chaotic changes in pressure and flow (see Figure 4-23 Figure 4-26 Figure 4-27). Turbulence is extremely complex; it is considered by many leading physicists as the most important unsolved problem of classical physics.

In the 19th century physicists like Stoke and Reynolds studied turbulence and it was found that laminar flow changed to turbulent flow when a parameter called Reynolds number exceeded a certain critical number. The Reynolds number is defined as:

	$Re = \frac{\rho \bar{u}_x D}{\eta}$	4.9-1
--	--------------------------------------	-------

The parameter D is a length scale, in the case of tube flow it is the diameter of the tube. The other symbols have its usual meaning of density, average velocity in the axial direction and viscosity. The Reynolds number is a dimensionless number which is a ratio between inertial and viscous forces in the liquid. The transition from laminar to turbulent tube flow occurs when Re is in the range between 2300 and 2600. The critical Reynold's number depends on geometry and properties of the tube wall. The critical number should not be taken too literally but rather as a rough indication for the transition.

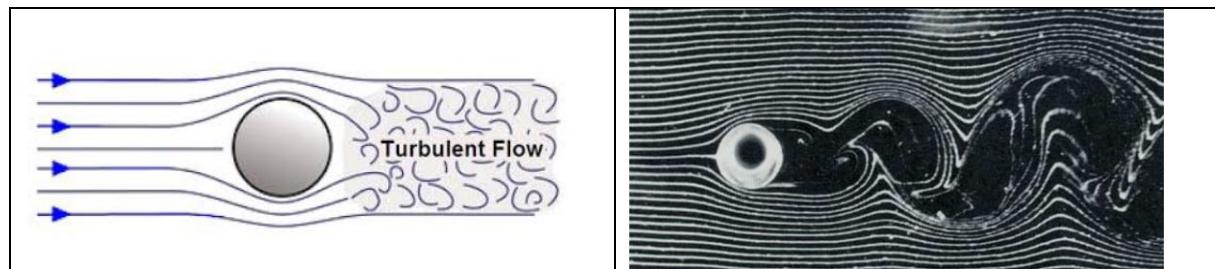
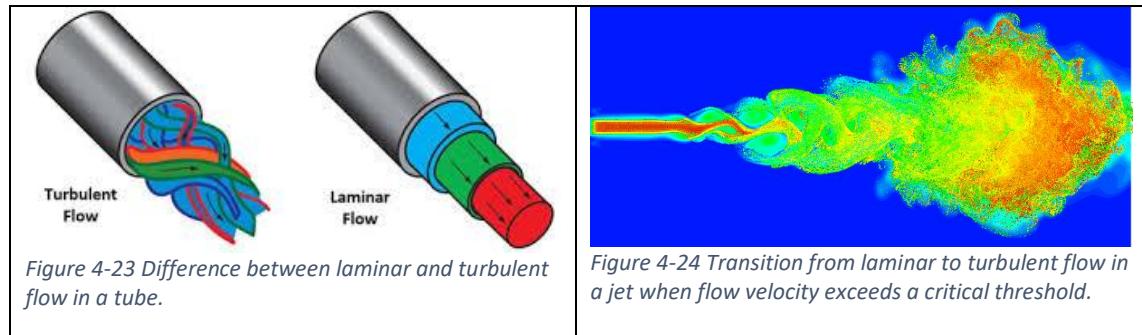


Figure 4-25 Left: Drawing of laminar and turbulent flow patterns for air flow around a cylinder. Laminar flow changes to turbulent flow after the cylinder. Right: Experimental visualization of flow lines around a cylinder

When flow is turbulent there are many vortices and eddies and many fluid layers with different velocities come in contact (Figure 4-25), as a result the viscous shear forces increase when compared with the laminar condition. This leads to higher energy dissipation in the fluid. Hence a larger pressure gradient is required to increase flow. This is shown in Figure 4-26 where the log-log plot of pressure gradient versus flow is shown. In the laminar region flow is proportional to the pressure gradient (Poiseuille). In the turbulent region flow is proportional to the square of the pressure gradient.

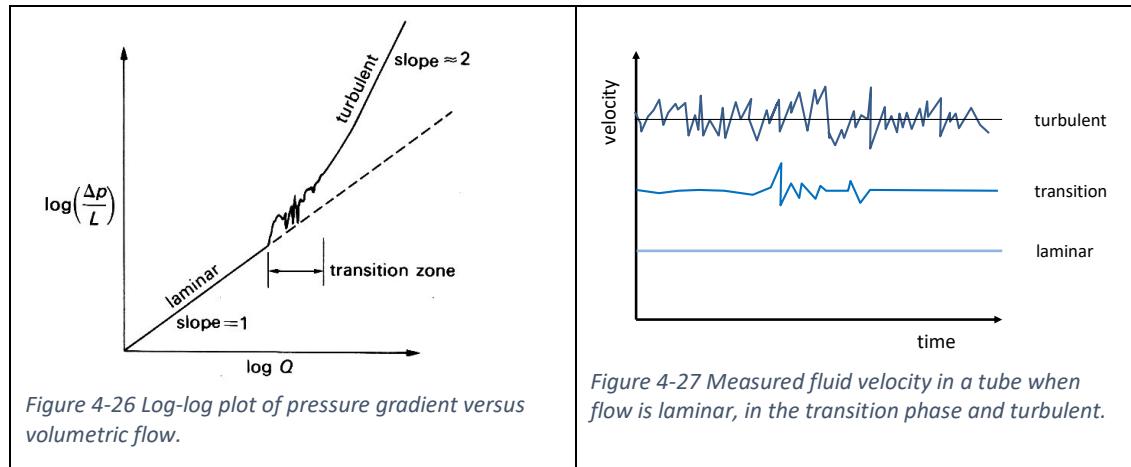
The resistance of a tube with turbulent flow can be approximated by the following relation:

$$R = R_0 + K \cdot Q \quad 4.9-2$$

For small volumetric flow Q the resistance is equal to that of the laminar flow value, for higher flows the resistance is proportional to the flow Q . The constant K is the Rohrer constant. The resistance increases when the flow becomes turbulent. This increase in flow resistance is important for larger air flow as occurs for instance in the large airways and in the tubing of the breathing set of mechanical ventilation machines.

Turbulent flow causes chaotic variations in flow velocity. In Figure 4-27 the flow velocity is shown versus time for three conditions (laminar, transition and turbulent). For the laminar flow velocity is

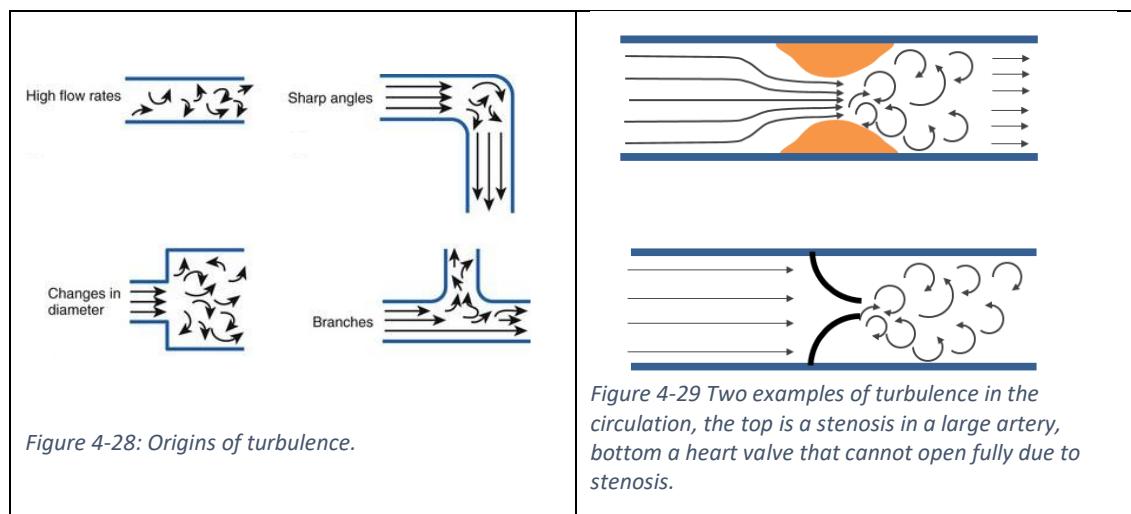
constant, in the transition region sometimes irregularities in flow are observed. For the turbulent type of flow there is a “noise like” variation in flow velocity with time. Turbulent fluid flow emits sound waves with a frequency in the audible range. With a sensitive microphone or stethoscope sound and the characteristics of the sound can be heard. Sounds generated by turbulent flow have a distinctive characteristic and can easily be recognized by a trained observer. In both respiration and circulation turbulence can occur and the characteristic sounds can be used by clinicians for diagnosis purposes. This will be discussed in more detail in the clinical measurement section.



Turbulence can occur in straight pipes but occurs earlier when local flow velocities change rapidly due to geometry effects. This is illustrated in Figure 4-25 and in Figure 4-28 where turbulence originates near sharp angles, a sudden change in diameter or a branches or bifurcations.

Physiological relevance

Such structures occur frequently in the respiratory and circulation systems. Two clinical examples of turbulence caused by stenosis of a large artery and stenosis of a heart valve are shown in Figure 4-29. Both lead to local reduced area for flow and flow velocity increases in the narrow opening. The flow velocity increases, the critical Reynolds number is exceeded, and turbulence occurs at the downstream side of the blockade.



Turbulence is often a very important part of the design of systems, again consider the nozzle design that was described in the previous section. The engineer has to verify if turbulence occurred, if this

occurs the calculation is valid and has to be modified. The engineer uses equation 4.9-1 to calculate the Reynolds number. His design was based on 10 cm inlet diameter and 1 cm outlet diameter. Using available data for air he finds $Re \approx 50000$ which is much larger than expected, severe turbulence occurs, and he has to redesign the system.

It appeared that the basic first order equations between volumetric flow and pressure gradients in fluid dynamics are like those of charge flow in the electrical domain. It is therefore possible to use electrical network analysis and simulation tools to model fluid flow in tubes. It is then possible to exploit the know-how and experience of the electrical engineering community. This is an enormous advantage. This approach will be used extensively in the rest of the course. The basic theory of electrical network analysis that is needed for the remainder of the text is recapitulated in the next section.

4.10 Tube network models

In this course extensive use is made of lumped element models for the circulation and respiration systems and sensing systems. These simple models are well suited for first order modeling of flows and pressures in physiology and for sensor transfer functions. Their simplicity is also attractive for understanding the main properties of physiological systems and give insight in causes and treatment of pathophysiology. Furthermore many sensing systems can be analyzed and modeled. A further advantage is the analogy of these models with networks in the electrical domain. The whole machinery in network analysis, equation solving, and simulation tools can be reused for physiology purposes. This is a huge advantage that is exploited in this course in all aspects. First the analogy between fluid dynamic tube flow and electrical currents and voltages is discussed followed by discussion of important formulas, network analysis methods and application to a simple network. More first and second order networks are shown appendix 13.2

4.10.1 Analogy between Fluid and Electrical Systems

The correspondence between the main parameters for fluid and electrical domains are listed in Table 4-1. The following pairs are linked: pressure and voltage, pressure gradient and electric field, volume and charge, flow resistance and electrical resistance, compliance and capacitance and inertance and inductance.

Table 4-1 Correspondence between fluid and electrical flow parameters.

Fluid transport	Electrical transport
Pressure P (Pa, scalar)	Voltage V (Volt, scalar)
Pressure gradient (Pa/m , ∇P , vector)	Electric field (V/m $E = \nabla V$, vector)
Volume V (m^3)	Charge q (Coulomb)
Volumetric flow Q (m^3/s)	Current i or I (Ampere, C/s)
Flow resistance R (Pa.s/m^3)	Electrical resistance R (Ohm, V/A)
Compliance C ($C = \frac{dV}{dP}$) (m^3/Pa)	Capacitance C ($C = \frac{dq}{dv}$) (C/V, Farad)
Inertance L (mass, $\text{Pa.s}^2/\text{m}^3$)	Inductance L (V.s/A, Henry)

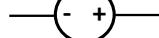
The basic equations for flow are listed in Table 4-2. The equations are similar. Note that there are subtle differences. For instance, in the electrical domain there is positive and negative charge, there is no such thing as negative volume. Furthermore, the scaling of resistance with radius differs a lot and this will have major consequences in the field of physiology. The concepts of laminar flow and turbulence do not occur in electric flow. The impact of gravitational forces is also negligible. The

approximations and limitations of techniques from the electrical domain are minor compared to the advantages. We will make extensive use of models and techniques from electrical domain to model fluid flow in the physiological domain.

Table 4-2 Basic equations for flows and currents in the fluid and electrical domain.

Fluid transport	Electrical transport
Resistor - Poiseuille Law $\Delta P = Q \cdot R, R = \frac{8\eta L}{\pi r^4}$	Resistor - Ohm's Law $V = I \cdot R, R = \frac{\rho L}{\pi r^2}$
Compliance $Q = C \frac{d\Delta P}{dt}$	Capacitor $I = C \frac{dV}{dt}$
Inertance $\Delta P = L \frac{dQ}{dt}$	Inductor $\Delta V = L \frac{dI}{dx}$

The following symbols will be used for the main lumped elements.

Flow resistor	
Compliance	
Inertance	
Pressure source	
Flow source	

When a lumped element or a continuum model is the most appropriate model for the specific use case is discussed next.

4.10.2 Selection of model elements

Models must be adapted to the degree of accuracy required and have to include the main physical effects. The first step is to check if a continuum model or a lumped model is needed. Pressure and flow in elastic tubes exhibit wave-like propagation. The wave type behavior can be neglected if the tube length is short with respect to the wavelength of the (pressure) pulse. The wavelength (λ) - frequency (f) relation for the fluid domain it is given by:

$$\lambda \cdot f = PWV \quad 4.10-1$$

As a rule of the thumb the long tube continuum model must be used when $L > \lambda/10$. Otherwise a lumped element model suffices (see Figure 4-30).

Physiological Relevance

For an artery a typical value for PWV is in the range 3m/s - 10 m/s. Fourier analysis of the waveforms show that the most important harmonics are in the range between 1 and 10 Hz. There the relevant wavelength is in the range of a few meters. The large artery system is longer than 1 meter, i.e. larger than $\lambda/10$, therefore a continuum or transmission line model is most appropriate for the arterial tree.

In this course lumped element models can be used in most cases. Even a continuum model can be modeled using a series of lumped element blocks. Lumped models simplify the description of distributed systems when wave phenomena are not important. It is desirable that the model should be as simple as possible, i.e. contain the minimum number of elements but still mimic the specific behavior of the system (see Figure 4-30). For instance, steady flow in a rigid tube can be modelled by a single flow resistor (1). Time dependent flow in a small radius elastic tube can be modeled by a

model of a resistor and compliance (2). Time dependent flow in an elastic tube with fluid viscosity and inertance included can be modeled with a three-element model (3) whereas unsteady flow in a long tube where distributed effects are of importance can be modeled by a series of the three element blocks (4), often five to ten blocks are sufficient for a good model.

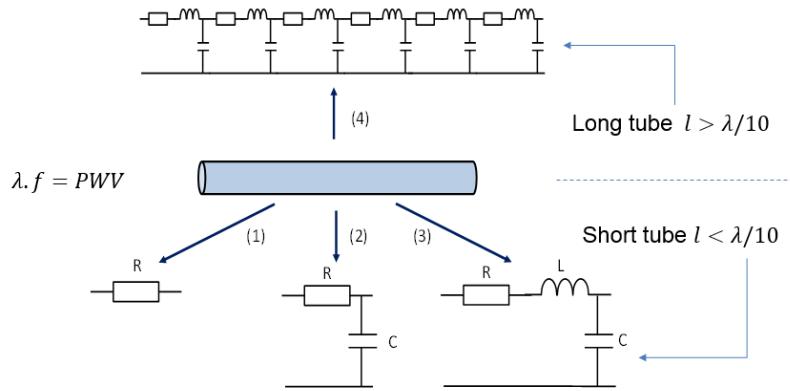


Figure 4-30 Lumped element tube models of different level of complexity.

The number of elements per section depends on the relative importance of resistance, compliance and inertance. In most cases compliance is important, however in some cases inertance dominates over resistance or sometimes it is the other way around. The relative important of resistance and inertance can be quantified by the ratio of the impedances of resistance ($Z_R = R$) and inertance ($Z_L = \omega L$). Using the equations 4.4-4 for the resistance R and 4.5-2 for the inertance L the ratio is equal to:

$$\frac{Z_R}{Z_L} = \frac{R}{\omega L} = \frac{8\eta}{\omega \rho r_0^2} \quad 4.10-1$$

Typically for small tube radius resistance (< 0.1 mm) resistance dominates, for large tube radius inertance dominates. Furthermore, the frequency is of importance.

Physiological relevance

Estimates for dominance of either resistance or inertance for arteries in the human circulation are shown in Figure 4-31. Average blood and arterial parameters are used. It appears that for large diameter arteries inertance dominates over resistance and resistance can be neglected. However, these large arteries are part of the arterial tree where a continuum TLM like model would be most appropriate. A lossless TLM model should give a good first order description of the pressure and flow waves in the arterial tree. For small diameter arteries and arterioles close to the tissues and cells the resistance dominates, inertance can be neglected. Since compliance is also important a simple two element RC circuit would suffice. Finally, the density (air versus blood) is important factor. In general, for air filled tubes and low frequency signals (< 1 Hz) inertance can be neglected even for the largest airways.

Blood parameters & arterial parameters

$\eta = 0.003 \text{ Pa.s}$
 $\rho = 1060 \text{ kg/m}^3$
 $\omega = 24 \text{ rad/s}$
 $PWV = 7 \text{ m/s}$



Radius (cm)	$\frac{R}{\omega L}$
1	0.01 ($<< 1$)
0.1	≈ 1
0.01	100 ($>> 1$)

Figure 4-31 Estimates of the ratio of resistive and inertance impedances for blood.

Finally, different forms of lumped models can be used for the same vessel segments. This is illustrated in Figure 4-32 for different embodiments of a three element RLC segment. Although the DC transfer function is the same and input and output flows are conserved the difference in input and output impedance can make a difference in model results and/or simulation stability and convergence.

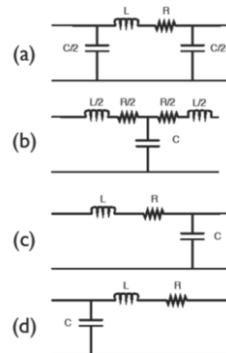


Figure 4-32 Different options for a three element RLC vessel segment.

Lumped element circuit models are used frequently in the cardiovascular and respiration systems. In this course these are used for both physiological and sensor systems. A sort summary of the most important network relations and methods to solve equations is discussed in appendix 13.2. As example a simple series and parallel RC and RCL networks are discussed in appendix 13.2.

4.11 References

1. Fluid Mechanics 8th edition by Frank M. White (Mc Graw Hill India, ISBN-10: 9385965492)

4.12 Questions

1. Which assumptions must hold if we are to assume that the flow through a tube has a parabolic flow profile?
2. Why a pulsatile flow is often assumed to have a "plug flow" profile?
3. What is the physical significance of the L, R and C elements when we model flow through a tube as in a long tube?
4. Calculate fluid inertance L and fluid resistance R for a tube of length 1 m and radius 3 cm (assume it is filled with water).
5. How do the values of L and R change when we replace the liquid (water) in a tube with a gas (air)?
6. How do the values of L, R and C of a fluid-filled tube change when a tube's radius doubles? And when its length doubles?
7. The tube in exercise 3 now has an elastic outer wall with young modules $E=500000 \text{ N/m}^2$ and a wall thickness of 1 mm. Calculate the tube compliance for a tube of a length of 1 meter.
8. Calculate the characteristic impedance Z_0 and pulse wave velocity of the elastic tube from the previous section (assume that it is lossless).
9. What is the hydrostatic pressure difference between head and feet if a patient a) stands up and b) lies down? (assume a head to feet distance of 1.7 m)
10. Explain how an external pressure can influence the flow through a collapsible tube.
11. Discuss the general characteristics of a second order transfer function. What are the natural frequency and the damping ratio? When can amplification occur for certain frequencies? Under which conditions is distortion minimal?
12. Explain the differences in input impedance of the different RLC segments shown in Figure 4-32.
13. For the cardio-vascular medical domain a doctor uses units of pressure and flow expressed in mmHg and l/s. Derive the units for resistance, compliance and inertance.
14. An ideal flow source with a step flow pulse of 800 ml/s and 0.1 second duration is applied at the input of a parallel RC segment. The resistance is 1.25 mmHg.s/ml, the compliance is 2 mmHg/ml.. Sketch the output pressure and flow through the resistor versus time.

5 Sensors, Transducers and Measurement Modules

A patient monitor has the function of a sensing device during clinical treatment. In the monitor vital-sign signals are measured, processed and presented to the clinician for further treatment and therapy. The monitor itself is a system that consists of a sensor, front-end electronics and a back-end system. In this chapter the sensors, transducers and basic measurements (pressures, flows, ECG, volumes (plethysmography), temperature, motion, ..) that are used in modern patient monitors are discussed. In the first section an introduction of sensors and transducers is given, the following sections will deal with sensors for pressure, flow, temperature, motion, electrical potentials, volume and gas-composition. Sensors that are used in state-of-the-art monitors are described. A more detailed description of these sensors can be found in the book of Webster [1]. Some sections in this reader are copied from the book of Blom [2].

5.1 Sensors and Transducers

A sensor is a device or subsystem that detects events or changes in physical, chemical or biological conditions of the environment. A sensor detects a physical or chemical quantity and converts the data into an electrical signal that is sent to other electronic systems such as a monitor or computer system. A transducer is a device that converts one form of energy to another form of energy, the other form does not need to be electrical. A sensor is also a transducer. In the remainder of the chapter the term transducer or sensor is used interchangeably. There are transducers that convert a difficult to measure parameter into an easier to measure parameter (for instance flow to pressure or temperature). The output of the transducer is converted to meaningful information for the user. There are passive transducers which require an external power source (e.g. a thermistor to measure temperature, a potentiometer) and active transducers that do not require external power (e.g. a thermocouple to measure temperature, photovoltaic cell). A diagram of a transducer system is shown in Figure 5-1.

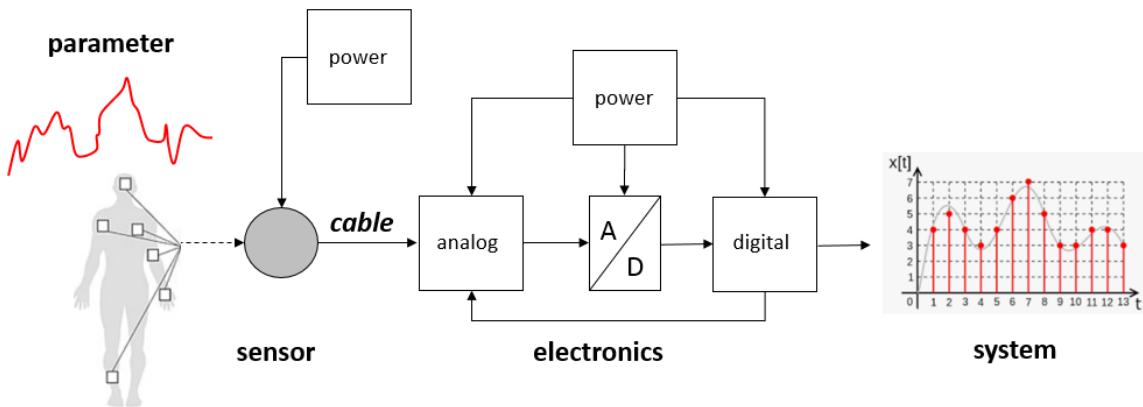


Figure 5-1 Diagram of a patient monitor transducer system.

In this figure a sensor is shown which is connected to a patient. The sensor measures a certain parameter and converts it to an electrical signal that consists of a DC baseline signal, the distorted signal, offsets, noise, interference and artifacts. The DC, artifacts and interference are often much larger than the signal of interest. Furthermore, the signal-to-noise ratio can be small. The sensor signal is transmitted via a cable to the analog front-end electronics that filters (removes part of the noise, limits bandwidth), amplifies the signal and converts it to the digital domain. The filtered and amplified analog signal is filtered by an anti-aliasing filter and is input for an analog-digital converter with a given sampling frequency. The sampling frequency depends on the required bandwidth of the signal. It is typically in the order of 100Hz, except for ECG where a sampling rate of around 500Hz to

a few kHz is needed. The AD converter sends the digitized time discrete signal to a microcontroller where further digital signal processing, digital filtering, noise and artifact removal, further digital processing such as demodulation and feature extraction is done (vital sign parameters). The processed signal (waveform) and extracted parameters can be sent directly to the host system where meaningful data are presented to the user. Note that due to the high level of analog and digital signal processing the waveform shown on the display may deviate from the ideal physiological waveform.

The sensor transfer function needs to be optimized and there are many constraints. The following list of sensor and system parameters need to be considered for a given clinical use case.

1. Sensitivity (output change per unit change of the input parameter)
2. Low sensitivity to other parameters (i.e. temperature drift, ...)
3. Large dynamic range (the ratio of the largest and smallest signal that must be measured)
4. Signal to noise ratio (including frequency dependence)
5. Accuracy (percentage deviation of the mean from the reference value)
6. Resolution (smallest change that can be measured with confidence)
7. Precision (variation in output signal for a fixed input signal, i.e. standard deviation)
8. Linearity and distortion (faithful reproduction of signal)
9. Bandwidth (relevant frequency range of signal)
10. Hysteresis (lag in time between signal and effect causing it)
11. Disturbance of the system to be measured
12. Artifacts and errors (i.e. motion, temperature, position ...)
13. Biocompatibility (i.e. not harmful for the patient)
14. Size, weight and cost
15. Power consumption (for battery powered use case)
16. Safety for the patient and user during single fault conditions
17. Reliability

The accuracy and precision parameters are illustrated in Figure 5-2.

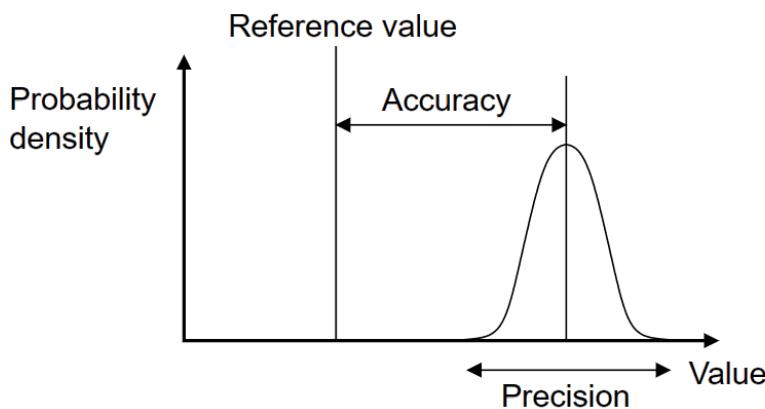


Figure 5-2 Diagram for illustration of accuracy and precision.

The order of the sensor system is an important parameter for the system design. The order is related to the order of the differential equation that describes the relation between input and output signals. These equations have been treated in previous lectures of elasticity and fluid transport. There are zero, first and second order systems. These are illustrated using the following diagram (Figure 5-3) that contains examples, equations and important parameters.

An ideal elastic spring or an ideal potentiometer are examples of zero order systems (frictionless and massless systems). The relation between input and output is a simple algebraic relation. The output signal is directly proportional to the position of the ruler, the gain K is given by the ratio of b_0/a_0 , there is no delay or phase shift between the input and output signals, these sensors are typically used in very low frequency applications (close to DC).

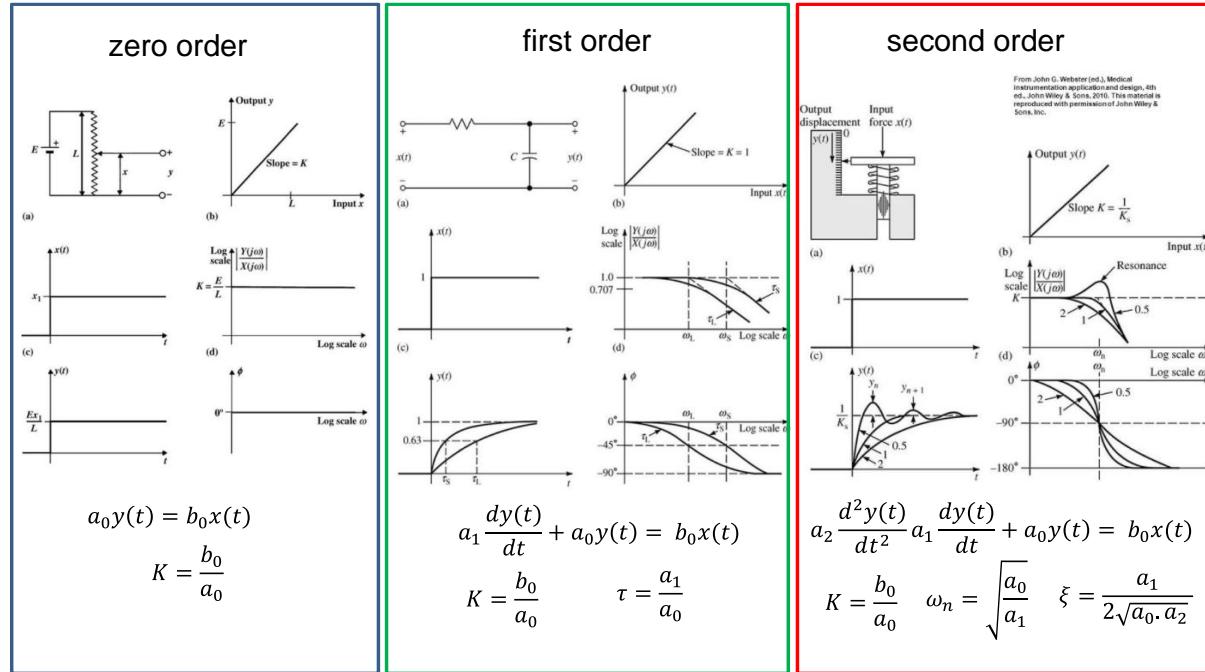


Figure 5-3 Diagram to illustrate the order of a transducer system.

A first order system is characterized by a first order differential equation. As an example, a simple low pass RC filter is used. The DC gain K and time constant τ are shown in the middle figure. For a step input in the time domain the output signal is an exponential function with a maximum signal equal to the input signal magnitude (DC gain K=1). The time constant determines the bandwidth of the filter. The frequency content of the output signal is limited. The output signal is always smaller than the input signal. There are no oscillations in the output signal. A temperature sensor is a first order sensor system. A potentiometer with friction between the sliding contact and resistor material is another example.

A second order system is described by a second order differential equation. As an example, a scale for weight measurement is used (e.g. the mass-damper-spring system described in appendix 12.3). Depending on the values of the a-parameters the system is damped or oscillatory. The DC gain of the system is similar to that of the first order system, but the frequency dependence is more complex. It is determined by the resonance frequency and damping parameter (ξ). The system should be designed such that oscillations and signal overshoots are limited. This can be done by designing for a damping parameter smaller than one, an optimum value is 0.7 (tradeoff between gain and phase shifts). When the damping parameter is larger than one, oscillations occur, and severe signal distortion occurs. A catheter/transducer system used for invasive blood pressure is a second order system. It is a fluid filled catheter tube, the combined inertance of the fluid and elasticity of the tube

wall and sensor can cause severe ringing and signal distortion and can result in large errors of the measured systolic and diastolic pressure.

5.2 Clinical Measurement

A clinical measurement⁹ includes the complete chain from the physiological signal to the extraction of processed waveforms, features and parameters that are used in the back-end of the patient monitor. This is shown in Figure 5-4. Hence a clinical measurement is much more than the sensor and analog front-end electronics.

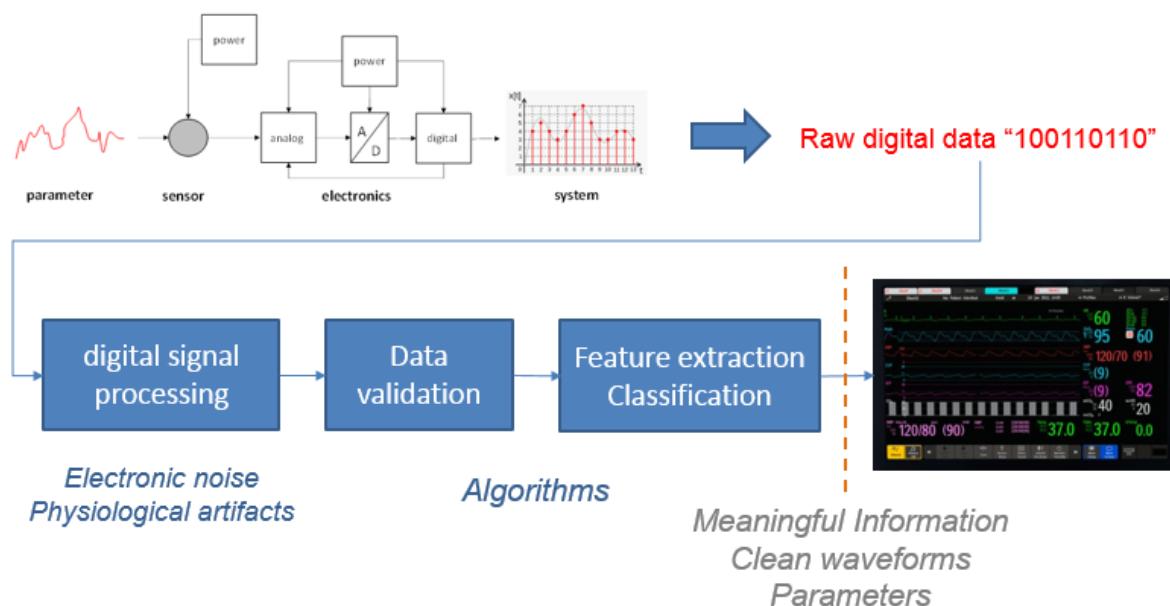


Figure 5-4 Diagram that shows the different parts of a clinical measurement.

A clinical measurement includes all stages from sensors, cables and connectors, device protection, signal sensing and acquisition, analog and digital processing and signal conditioning, adaptivity, removal of artifacts, noise reduction, validation of the data and the measurement, error reporting, and thereafter feature extraction and parameter extraction. Quality can be guaranteed best when all parts of the measurement are developed by the same organization. For instance, waveforms shown in the monitor can be affected by signal processing and filtering in the measurement module. Certain features are filtered out and, removed and cannot be detected afterwards. *Hence a clinical measurement is defined in this course as the complete chain from signal sensing down to the generation of meaningful information to the monitor.* The patient monitor will add other functions and processing such as display and alarm generation. In some cases much more processing power is needed for feature extraction, an example is the detection of a heart rhythm from ECG measurements. In this case the waveform from the measurement is further analyzed in the host system where much more processing power is available.

In the following chapters transducers for pressure, flow, volume, temperature and motion are described. Thereafter three clinical measurements of vital sign parameters are described (ECG, Capnography and Pulse Oximetry).

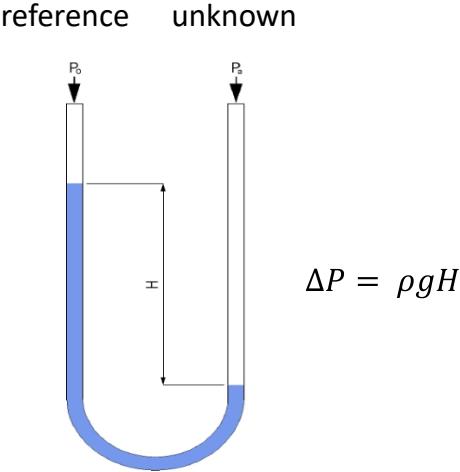
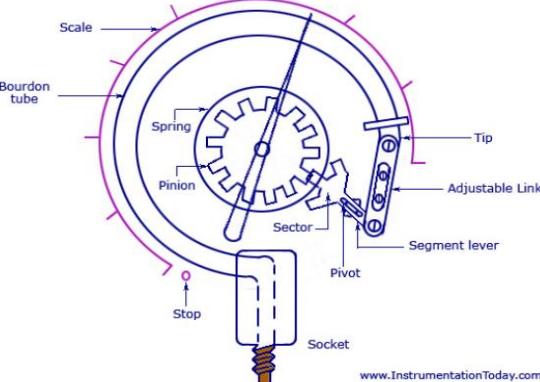
⁹ Standard used in Philips monitors

5.3 Pressure Measurement

Pressure can be measured with high accuracy and precision. Resolution is often much better than needed for the measurement. There are many pressure transducers, both of a mechanical (active transducer) and electronic type.

5.3.1 Mechanical Transducers

Two pressure transducers used for manual blood pressure measurements are shown in Figure 5-5 and Figure 5-6. In Figure 5-5 a fluid filled manometer is shown. The pressure difference is proportional to the difference in the height H of the fluid meniscus between the tubes. The accuracy ($\sim 1\text{mmHg}$) and precision of this device are excellent, for a long time this device was the gold standard for clinical use. For blood pressure measurements mercury filled manometers have been used extensively but are phased out because of safety reasons (toxicity of mercury).

 $\Delta P = \rho g H$	 <p>Figure 5-6 Drawing of a Bourdon type aneroid sphygmomanometer.</p>
 <p>Figure 5-7 Mercury manometer for a cuff based blood pressure measurement.</p>	 <p>Figure 5-8 Photo of an aneroid sphygmomanometer for a cuff based blood pressure measurement.</p>

A Bourdon type mechanical transducer is now more common and is shown in Figure 5-6. A pressure change in the flattened Bourdon tube changes the tube cross section and the tube stretches as the pressure increases. This length change of the tube is converted via a mechanical gear transmission to a pointer. The pointer position reflects the pressure which is read from a scale. The embodiments for a cuff based non-invasive blood pressure measurement of the two transducers are shown in Figure 5-7 and Figure 5-8. The Bourdon manometer is still used extensively for non-invasive blood pressure

measurements. It is the gold standard for mechanical non-invasive blood pressure measurement (accuracy ~1% of full scale). The mechanical transducers are not suited for automated measurements. Electrical sensors are used for monitoring devices and are described in the next section.

5.3.2 Electronic Pressure Transducers

The bending of a flexible membrane due to a pressure gradient over the membrane causes a change in values of a resistor or capacitor element that is placed on the membrane. Changes in the parameters of these sensing elements (dimensions, spacing, specific resistivity) are used to measure a differential pressure. This is shown schematically in the diagram of Figure 5-9. When a differential pressure is applied the thin membrane bends or deforms. The bending is inward when the differential pressure is negative and outwards when the differential pressure is positive. The bending of the membrane can be measured in several ways. The two most used techniques for pressure measurement are described below.

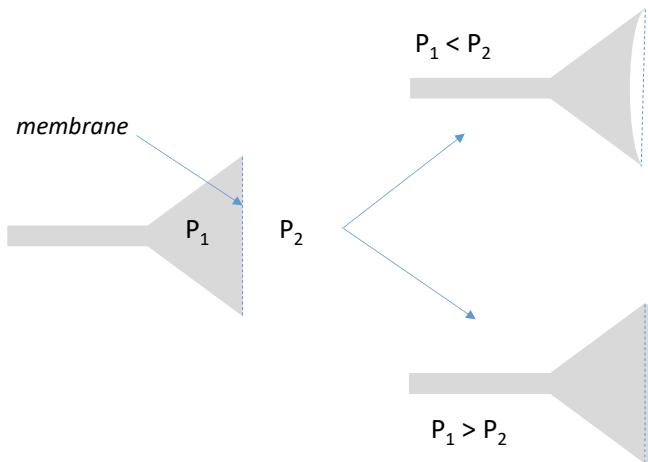


Figure 5-9 Diagram of an electronic pressure sensor based on membrane deflection. The original position of the membrane is indicated by the blue dashed line.

They are shown in the diagrams of Figure 5-10 and Figure 5-11. The solution which is used most often in the clinical domain is the strain gage. Piezo resistors (change in resistance due to strain) are integrated in or placed on top of the foil (see Figure 5-10). The resistance varies when the resistor is deformed mechanically. The change in resistance is in first order proportional to the change in pressure. Changes in resistance can be measured with high accuracy and precision (<< 0.1%). Furthermore, by design and material selection the effects of drift and temperature changes can be reduced.

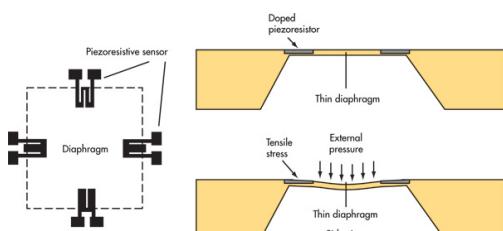


Figure 5-10 Diagram of a strain gage pressure sensor.

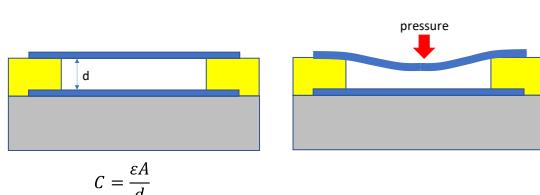


Figure 5-11 Diagram of a capacitive pressure sensor

The second method is a capacitive method. When the foil bends, the separation between the foil and a reference electrode changes (see Figure 5-11) and a change in capacitance occurs. This change is inversely proportional to the separation between the two plates. Changes in capacitance can also be measured with very high accuracy and precision using bridge techniques. Very small changes in pressure can be measured (fraction of a mmHg). The cost of the capacitive method is normally much

higher than that of the strain gage solution. However recently fully integrated transducers in (Bi)CMOS IC processes have become available and cost of this solution is now low enough for application in lifestyle devices (mobile phone, sport watches) for measurement of the barometric pressure (measurement of changes in altitude).



Figure 5-12 Commercial strain gage pressure transducer for professional use from Honeywell. The integrated solution is shown on the left, a MEMS sensing device is shown on the right.

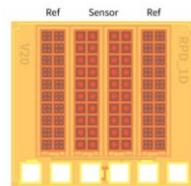
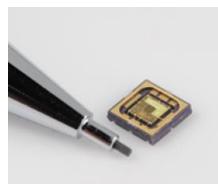
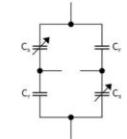


Figure 5-13 Capacitive barometric pressure transducer from Infineon developed for lifestyle applications.



Two commercial solutions for differential pressure sensing are shown in Figure 5-12 and Figure 5-13. A complete transducer with connections for gas pressure sensing and digital electrical output signal from Honeywell are shown in Figure 5-12. The sensor is fabricated using MEMS technology, the sensor and electronics are in the same IC package. For this specific sensor the accuracy specification is 0.75 mmHg and the resolution is less than 0.1 mmHg (13 Pa). For gas pressure sensing the bandwidth is around 1 kHz. This type of sensor is well suited for non-invasive blood pressure measurement techniques. A capacitive sensor from Infineon is shown in Figure 5-13. The transducer is a fully integrated solution, the capacitors and electronics are on the same die. This enables a very high resolution of 0.5 Pa (0.004 mmHg). This transducer was developed for sport watches and smart devices for barometric pressure sensing to measure height differences during walking, detection of using a staircase and for outdoor sporting. A change in height of only 5 cm can be detected. This type of sensor was used for fall detection in Philips Lifeline products. The clinical measurements of pressures in the respiratory system and cardiovascular system are discussed in more detail in lectures of module 4.

5.4 Flow Measurement

The measurement of flow is of great importance for clinicians. However, the measurement of gas and liquid flow is much more difficult than the pressure measurement. Often flow is converted in a pressure which is sensed. There are many options, each is optimal for a specific use case. The most used flow transducers in the respiratory and circulatory measurements are described below.

5.4.1 Pneumotachograph

A pneumotachograph is a device that uses the Poiseuille law to measure gas flow. Basically it is a flow resistor, the measured differential pressure is proportional to the volumetric flow rate. Note that this law is valid only for laminar flow. To avoid turbulent flow at high gas flow rates the diameter of the tube must be very small. The flow resistance would be unacceptably high for a single narrow tube. The obvious solution is to combine many small diameter tubes in parallel, this limits the flow resistance of the device. This principle is used in the pneumotachograph. A diagram of a transducer is shown in Figure 5-14.

The flow resistance of this device is reduced, a very sensitive and high-resolution differential pressure sensor is needed (range a few cmH₂O). The relation between pressure and flow is linear, it is an example of a zero-order transducer. The flow resistance depends also on the viscosity and temperature of the gas; a gas specific calibration is needed. The temperature of the device is often

raised above room temperature to eliminate temperature changes of the sensor and avoid condensation of water vapor in the expired air. The transducer is suited for higher flow rates (1 l/s) and can also handle bi-directional flows (inhalation and expiration). The volume of the sensor is a drawback, it increases the dead space in the system. Furthermore the pressure drop at high flow rates (60 l/min, 1l/s) is considerable, a pressure drop around 1.5 to 2.5 cmH₂O is typical, this in the order of 10% of the measured downstream pressure.

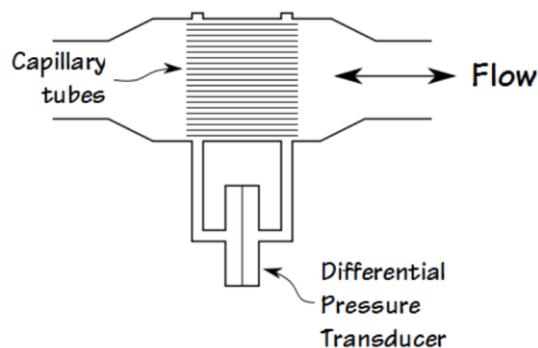


Figure 5-14 Diagram of a multi capillary pneumotachograph.

This pneumotachograph or variants of this device is often used as a gas flow sensor for patient monitoring and mechanical ventilator devices. The variants are designed for lower flow resistance. An example is shown below. Two disposable proximal flow sensors (i.e. air flow measured close to the airway opening of the patient) are shown in Figure 5-15. The Hamilton device consists of a thin diaphragm asymmetrically placed in the length direction in a tube of 15 mm inner diameter. The diagram bends when there is flow, this causes a change in resistance and the pressure difference between the two compartments is measured by a differential pressure sensor in the mechanical ventilator. The pressure difference is related to the flow rate. The device in the right is an integrated flow sensor, the upstream and downstream pressure and its difference is measured by a (CMOS) MEMS sensor with integrated electronics. The output flow is a digital signal sampled at 100Hz and is transmitted to the host device via I2C or SPI interface.



Figure 5-15: Two sensors used in mechanical ventilation devices, the Sensirion is an integrated device with digital output.

5.4.2 Venturi Flow Rate Transducer

Another category of flow sensor is based on Bernoulli's effect for flow in a pipe with two different diameters. It is widely used in engineering and clinical applications (ventilation machines, spirometers). A diagram is shown in Figure 5-16. The device is designed to minimize flow resistance and turbulence. The flow velocity can be obtained from Bernoulli's law by measuring the pressure difference $P_1 - P_2$ and using the known geometries and gas density. The volumetric flow can be measured by multiplying the flow velocity and tube diameter. The transfer function is indicated in the figure. The relation between differential pressure and flow velocity is not linear. Corrections for

resistive losses at high flow rates can be corrected in the electronics. An example for clinical use is shown below. It is the Philips Capnostat sensor that is used for simultaneous airflow and capnography measurements during mechanical ventilation. The capnography sensor is not shown.

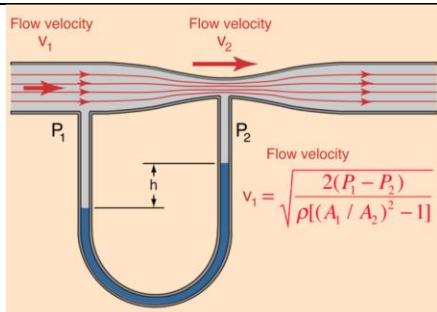


Figure 5-16 Diagram of a Venturi flow transducer.
The transfer function is indicated in the figure.

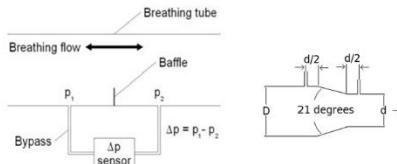


Figure 5-17 Diagram and picture of the Philips Capnostat flow and pressure sensor.

5.4.3 Pitot Tube Volumetric Flow Sensor

The Pitot tube is based on Bernoulli's law. The device is shown in Figure 5-18.

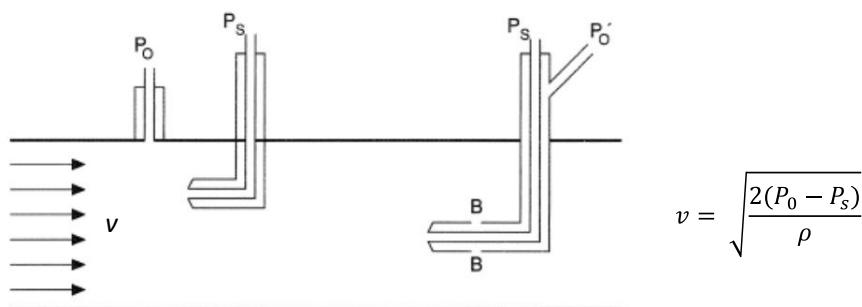


Figure 5-18 Diagram of two embodiments of a Pitot tube flow transducer. The transfer function is indicated in the figure.

The basic Pitot tube (labeled P_s) points directly into the gas flow, the pressure measured by the basic tube is the stagnation pressure P_s . Gas flow into the tube is blocked, the flow velocity is zero at the tube opening. The pressure at the opening is called the stagnation pressure and is equal to the internal pressure. The dynamic pressure is zero at the stagnation point and the hydrostatic pressure is a common mode pressure that is not measured by the differential pressure sensor. The other port labeled P_0 measures the total pressure that includes both internal and dynamic pressure (hydrostatic pressure can be neglected). Using Bernoulli's law the expression for the transfer function can be obtained, it is indicated in the figure. The relation between gas velocity and differential pressure is not linear. This can be corrected in the electronics. Pitot tubes are used to measure air speed in aviation and high-speed car racing. The principle is also of important for invasive blood pressure as it can introduce systematic errors in blood pressure, this will be discussed in module 4.

5.4.4 Rotameter

A rotameter consists of a slightly conical tube with a floating bobbin in it. A diagram of a rotameter is shown in Figure 5-19. Gas flows upwards and forces the bobbin in upwards direction. The bobbin has grooves on the side surface, it rotates around the central axis to keep it free from the tube wall. The position of the bobbin is determined by two forces that balance each other. The first force is a flow induced force which is proportional to the horizontal area S of the bobbin and the pressure difference over the bobbin. The opposing force is a gravity induced force which is equal to product of

bobbin mass and gravitational constant g . This gives an expression for the pressure difference ($\Delta P = mg/S$). When Poiseuille's law is valid a relation between ΔP and the gas flow can be derived, it depends on the area A , the difference between the local tube area and the area S , and the viscosity and composition of the gas. When the flow increases the bobbin is pushed upwards, the area A increases, flow resistance decreases until again the equilibrium between flow force and gravitation force is reached. A rotameter can only be used constant flows and for a certain gas (or gas mixture) at a specific pressure and temperature. A rotameter calibrated for oxygen flow is shown in Figure 5-19.

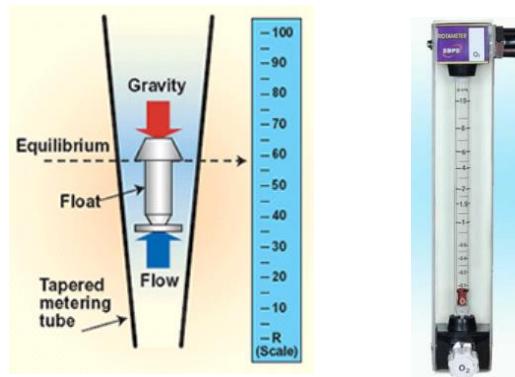


Figure 5-19 Diagram of a rotameter (left). Rotameter used for oxygen gas flow control (right).

5.4.5 Hot wire anemometer

In a hot wire anemometer a known current I is sent through a small resistive element which is immersed in the fluid flow (see Figure 5-20). The fluid flow will cool the element. The wire is made of a material whose resistance is temperature dependent. The hot wire resistance R , whose value V/I can be determined by measuring the voltage across it, will thus depend on the gas flow. The current I is made directly proportional to V ; this can be done by a feedback system. Thus, the value of R remains constant. A constant R implies a constant temperature T , effectively reducing the anemometer thermal time constant to a very low value. This produces a large bandwidth, so that rapid flow changes can be measured. A disadvantage, however, is that no information is available about the direction of the flow. The hot wire anemometer's reading depends on the amount of heat that is transferred from the wire to the gas, and this in turn depends on the temperature, pressure and composition (which determine density, viscosity, heat capacity and heat conduction coefficient) of the gas that flows around the sensing element. Because the hot wire can be made very small, the hot wire anemometer is an ideal instrument to probe the flow profile in a tube. For measuring the total flow its small size is a disadvantage, because errors will arise if the flow profile changes. This sensor measures the mass flow, it is a precursor of the mass flow sensors described in the next section.

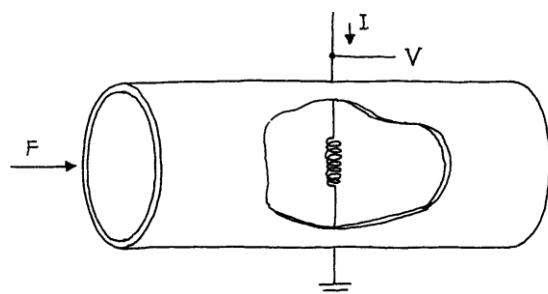
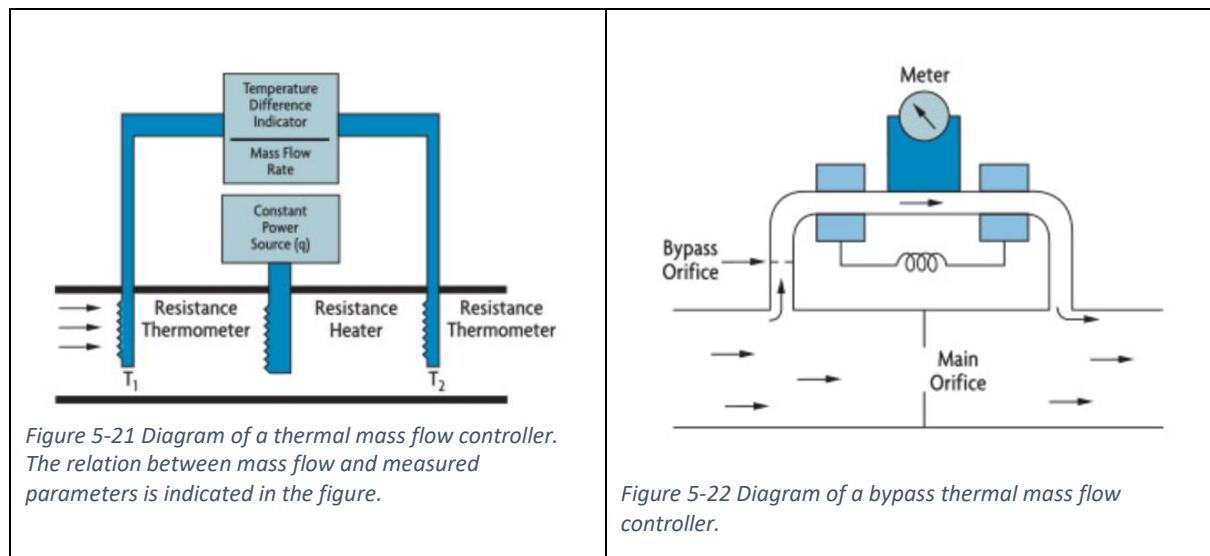


Figure 5-20 Diagram of a hot wire anemometer.

5.4.6 Thermal Mass Flow Transducer

The pneumotachograph and related devices described earlier measure the volumetric flow which depends on gas composition, pressure and temperature. The mass flow rate is an absolute measure of flow (i.e. kg/s or number of particles per second), it is directly related to the mass or number of molecules flowing per unit of time into the system. The two most used mass flow meters are of the Coriolis and thermal type. We will focus on the thermal type. A thermal mass flow meter measures the mass flow of gases and liquids directly. It is a variant of the hot-wire-anemometer. A diagram is shown in Figure 5-21.



The principle of operation is to transfer a known amount of heat per unit time (\dot{q}) to the fluid by convection of a heated element to the fluid and measure the temperature difference between the two temperature sensors in the upstream (T_1) and downstream direction (T_2). The mass flow rate is inversely proportional to the rise in temperature difference and specific heat C_p . The mass flow can be calculated using the relation:

$$\dot{m} = \frac{K \cdot \dot{q}}{C_p \cdot (T_2 - T_1)}$$

5.4-1

K is a calibration constant that is gas specific. For an ideal gas the measured mass flow does not depend on fluid composition, pressure and temperature. Immersion of the thermal heater element and the thermal sensors in the fluid can cause reliability problems and disturb flow. A more practical embodiment is a bypass type device that is shown in Figure 5-22. The heater elements are placed outside the bypass pipe improving reliability and enables measurement of large flows.

5.5 Blood Flow Measurement

Some of the transducers described in the previous section can be used for both gas and liquid flow. However, most of the devices cannot be used for blood flow measurements, it is not possible or allowed to place obstructions into in blood vessels. In this chapter we shall limit the discussion to transducers that can be used to measure blood flow in the clinic. These techniques are often limited to measure blood flow velocity rather than flow. In some cases flow can be computed if the velocity profile and lumen cross section can be measured. Some sensors are not used in measurements on patients but are used in some medical devices. The electromagnetic and ultrasound transit time

sensors are discussed in appendix 14.1 and 14.2. Measurements of the cardiac output are described in chapter XXX. The focus of this section is on ultrasound techniques.

5.5.1 Ultrasound Transducers

Ultrasound is used in clinical care for both imaging and for flow velocity measurements. Ultrasound devices measure flow velocity and not flow. The frequency range of interest is between 1 MHz and 10 MHz, the velocity of sound c is 1500 m/s. The wavelength of a 1 MHz sound wave is 1.5 mm. Ultrasound waves are reflected at sites where the acoustic impedance changes (material property, product of density and acoustic velocity). The reflected wave is in first instance used for imaging (B-mode ultrasound). When the wave is scattered by a moving object, the frequency of the scattered wave changes, this is the Doppler shift. The frequency shift is proportional to the velocity of the moving object. Red blood cells are very efficient scattering sites. Flow velocities of blood can therefore be measured with good accuracy. Volumetric flow can be estimated when flow profiles and tube dimensions are known.

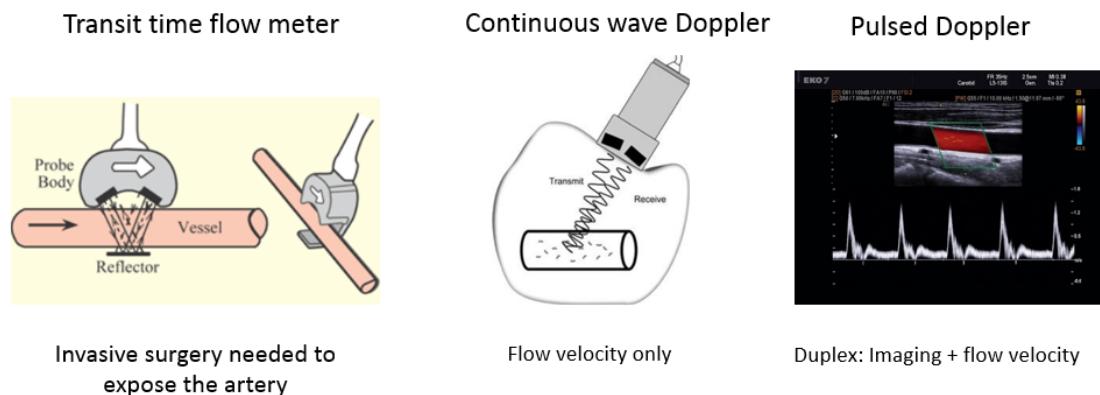


Figure 5-23 Ultrasound technologies to measure blood flow velocities.

Three methods are used to measure flow velocity of blood. They are illustrated in Figure 5-23. The first one is the transit time method; a double transducer and reflector are placed around a tube or blood vessel. The transmitter and receiver function is interchanged at a fixed frequency. The difference in transit time of the forward and reverse US is a measure of the flow velocity. It is a very accurate method but placement requires surgical procedures. Hence it is an invasive sensor and not suited for routine use. It is mostly used in animal models in a laboratory setting (it is described in appendix 14.2). The second method is continuous ultrasound. A transducer emits a continuous wave of ultrasound with a fixed frequency. The reflected wave is detected by a second transducer element and electronically processed to obtain the Doppler frequency shift and flow velocity. This measurement gives information on the average flow velocity in the cross section of a tube. The third and most versatile option is pulsed ultrasound. The transducer emits bursts of ultrasound pulses with a fixed repetition frequency. In the interval between the transmitted waves the same transducer is used to detect the reflected waves. The reflected waves are analyzed, an image is formed, and flow velocities are extracted from the Doppler-shifted part of the signal. This is the so-called duplex echography, both the image, region of interest and flow velocities as function of time are shown in a dedicated plot. It is now possible to measure flow velocities at various positions in the tube, i.e. a velocity profile and flow rate can be obtained

5.5.2 Transducer Elements

Ultrasound (US) transducers convert AC electric signals into soundwaves. There are piezoelectric (Figure 5-24) and capacitive transducers. A piezo transducer consists of a piezo crystal of diameter D between two metal electrodes, the thickness of the transducer depends on the electrical voltage over the piezo material. An AC voltage at the transducer resonant frequency is applied to the electrodes, the crystal oscillates with this fixed frequency and an ultrasound beam is formed. The transducer is designed for one specific frequency. The transducer can also function as a receiver, reflected soundwaves are converted into electric signals. The beam shape in the tissue is not constant (see Figure 5-25). The beam consists of a near-field part and diverging far-field part. The length of the near-field part depends on size of the transducer and wavelength.

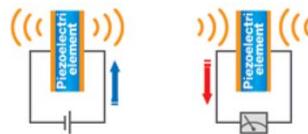


Figure 5-24 Piezo electric ultrasound transducers and a diagram of a transmitter and receiver.

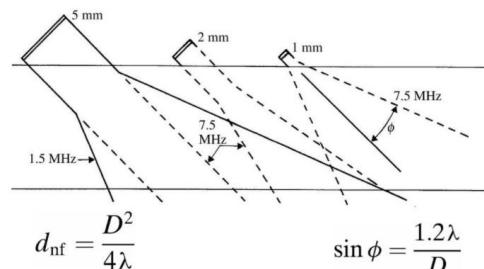


Figure 5-25 Near field and far field beams.

When high resolution imaging is needed the frequency must be increased and acoustic lenses are used to focus the diverging beam (see Figure 5-26). However, the absorption of ultrasound increases with frequency. Hence for structures deeper in the body high-frequency US cannot be used, the reflected signal amplitude is too small. As a result, measurements deeper in the body require low frequency ultrasound and resolution is limited. Transducers are optimized for a specific use case. A modern multi-element transducer suited for duplex ultrasound is shown in Figure 5-26. The design is optimized for optimal efficiency and includes focused beam formation and beam steering. Beam steering is done in a similar way as in a phased array radar system. In this way the beam position can be scanned, and images of a larger area can be created.

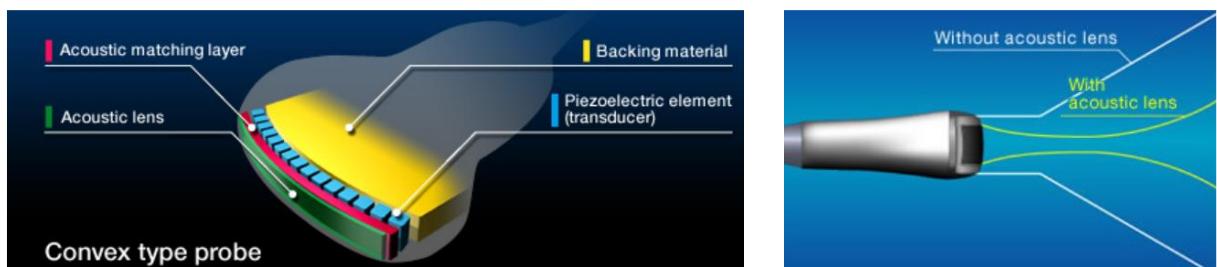


Figure 5-26 Linear array ultrasound transducer with acoustic matching and acoustic lens. The right picture shows the beam with or without the acoustic lens.

One of the drawbacks of a piezo transducer is the fixed frequency, fixed wavelength and fixed resolution. Furthermore, transducer and system cost are very high. The linear and two-dimensional multi-element transducers are very expensive which limits use. Using modern integrated circuit technology and MEMS technology, new very small and potentially low-cost transducers are possible. A micro-machined transducer (MUT) fabricated in a standard CMOS technology (CMUT) is shown in Figure 5-27. This technology can be used for both single and multi-element transducers that can function at a wide range of frequencies. Furthermore, miniaturized fully integrated transducers are possible as shown in this figure (device in catheter tip). Reliability and packaging of the transducer

are major challenges that still require further effort.

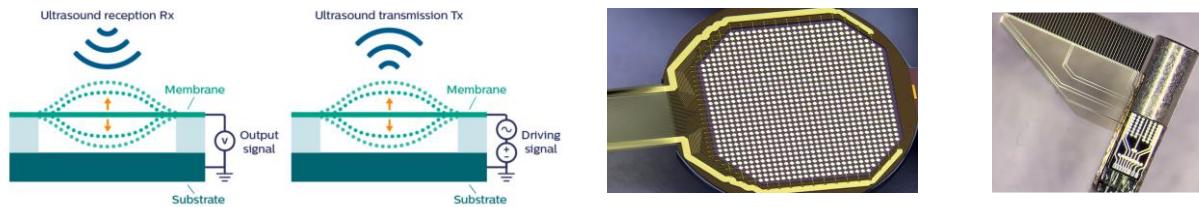


Figure 5-27 Diagram of a CMUT transducer, a two-dimensional array and a miniature transducer in a catheter tip.

5.5.3 Continuous Wave Doppler

A diagram of a continuous wave (CW) Doppler flow set up is shown in Figure 5-28. A transmitter element Tx is placed at an angle α with the axis of the blood vessel. The receiver element Rx is placed at an angle β with the axis of the blood vessel. The measured volume is formed by the overlap between the two beams.

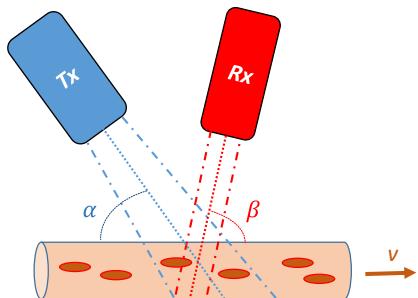


Figure 5-28 Diagram of a continuous wave Doppler flow set up with a transmitter Tx at an angle α with the axis of the vessel and a receiver element Rx with an angle β with the axis of the vessel. Transmitted and reflected beams are indicated.

The element Rx receives ultrasound scattered by red blood cells which move with velocities v , the velocity of the red blood cells varies over the measurement volume. The signal from the transmitter must be in a directional narrow beam to ensure that a single blood vessel is measured. The receiver element should receive signals from blood vessel segments with equal size over the diameter of the vessel. For a transmitted frequency f_0 the received frequency of a wave that is scattered by a red blood cell which moves with velocity v is equal to:

$$f = f_0(\cos\alpha + \cos\beta) \cdot v/c \quad 5.5-1$$

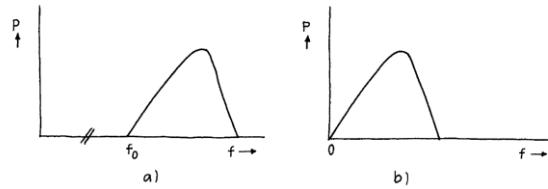


Figure 5-29 The received signal's frequencies vary around the transmitted frequency (a); a large echo at f_0 due to non-moving tissue is not shown. Synchronous demodulation subtracts the transmit frequency f_0 , resulting in a low frequency spectrum (b).

The frequency shift is proportional to the velocity of the red blood cells. The shift consists of two terms, the first is due to the velocity difference between the transmitter and the cell, the second shift occurs by the velocity difference of the transmitting cell and the receiver. The cosine terms are needed because of the difference in direction of the velocity of the cells (v) and the axis of the Rx and Tx devices. The Doppler shift depends strongly on the angles α and β . This can be a major source of error. The frequency f_0 is in the MHz range, the Doppler shift is in the order of a few kHz. In Figure 5-29 the received signal (a) and demodulated signal (b) are shown. The frequencies of the Doppler shifted signal are close to f_0 but synchronous demodulation removes the non-shifted high frequency signal from the received signal. When a quadrature demodulation technique is used the direction of flow is also measured.

The Doppler shifted signal spectrum can be visualized with the aid of a spectrum analyzer. The spectrum analyzer calculates the amplitude of the different frequencies of the demodulated signal

for specific time interval. In the spectral display (Figure 5-30) the frequency is on the vertical axis and the brightness of a pixel is related to the magnitude of the amplitude. The time is on the horizontal axis. Two examples of spectral Doppler are shown in Figure 5-30. Such plots give insight in changes of flow with time and on the distribution of flow velocities. There are clear differences between case a and b. In the former the velocity distribution is broad many velocities are observed, in the latter the flow velocity is more uniform, furthermore backflow occurs.

Frequency shift

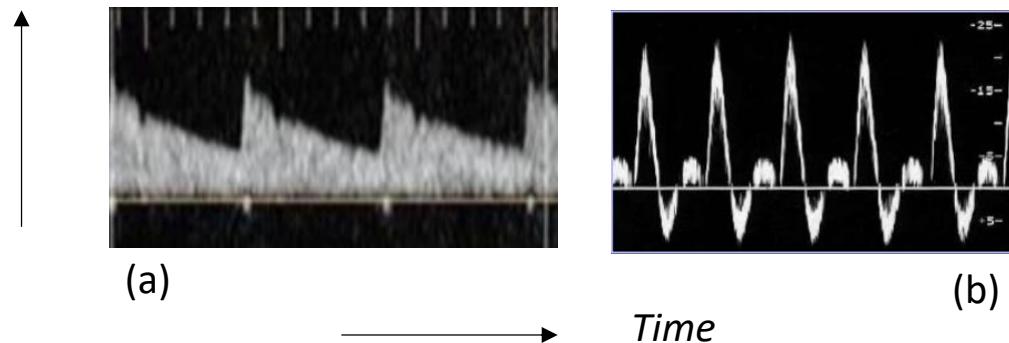


Figure 5-30: Two examples of spectral Doppler.

This technique measures blood velocity, not blood flow. In a few cases where the vessel's lumen diameter is known and there is plug flow, flows can be estimated. The technique is very useful, is non-invasive, patient friendly. It is used to detect the positions of obstructions in vessels; where a vessel is very narrow, e.g. due to an atherosclerotic plaque, high velocities will exist. It is a relatively low cost technology.

5.5.4 Pulsed Ultra Sound ([Blom])

Continuous-wave Doppler devices provide little information on flow profiles and often include signals from other blood vessels nearby. The pulsed ultrasound technique was developed to come up with a solution that solves the above problems. It offers many advantages; it includes both imaging and flow information and enables estimates of the velocity profile in large blood vessels. In the pulsed ultrasound technique, a single transducer is used for transmit and receive. A diagram is shown in Figure 5-31 (a).

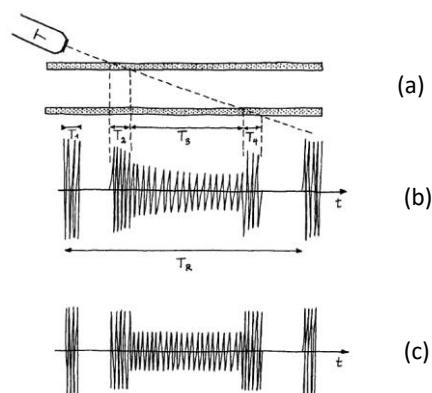


Figure 5-31 Echoes are generated when the beam of transmitter T encounters an acoustic impedance change (a). The time trace of the echoes is a one-dimensional "picture" of the tissues (b). Depth-dependent damping can be eliminated by a time-varying amplification (c).

The transducer, placed on the skin, transmits a short highly directional burst of ultrasound with a frequency of several MHz into the tissues of the body. The ultrasound is reflected wherever it encounters a change in acoustic impedance along its path. This echo can be directional if it is "mirrored" by a smooth surface (angle of reflection = angle of incoming beam), or it can be non-directional (scattering) by areas of differing acoustic impedance that are much smaller than the signal's wavelength. In biological tissues, no mirror-smooth surfaces exist; all echoes are (also) partly scatter reflections. In general, echoes will be received from all beam positions where a change in anatomical structure occurs. In this particular application, as indicated in Figure 5-31 (a,b) where the transmitted beam has been idealized as very narrow, echoes are received from the vessel walls (T2 and T4) and from blood cells (T3). T1 is the time during which the transmitter is active, and the next pulse is transmitted after the repetition time TR. Note that the power of the received signal decreases with the distance that the signal has traversed (back and forth); this is due to the absorption of ultrasound energy by the tissues. Since this absorption is approximately constant for all tissue types (except bones), it can be compensated for with a time-varying (exponentially increasing) amplification of the received signal (Figure 5-31 (c)).

The received signal is thus a one-dimensional "picture" of the anatomical structures that the transmitted beam traverses. Vessel walls are easy to recognize (T2 and T4). The signal that is received during T3 is due to the scattering by blood cells within the vessel. When these are in motion, the received signal is, due to the Doppler Effect, shifted in frequency. This part of the signal can be isolated by detecting when the large reflections due to the vessel walls stop (start of T2) and start again (end of T4) and multiplying the received signal by zero except during this time ("gating"). The average blood velocity can then be determined in the same manner as in the continuous ultrasound technique. What is new is that there is now also information about the positions of the vessel walls, which can be used to estimate the vessel's diameter and thus its total flow. When the vessel is the initial segment of the aorta, before it branches, this technique provides a continuous readout of the cardiac output. What is also new is that vessel wall motion can be followed by observing the varying positions of T2 and T4.

In order to compute the flow from average velocity and vessel diameter, all blood velocities must contribute equally to the received signal. If the beam is wide enough to enclose the whole vessel, this is ensured by processing the signal during the full period T3. Instead of this, we can also process only a very short segment of T3 ("gating") and shift the position of this segment between the end of T2 and the start of T4. If we do this, we obtain the blood velocity at a specific depth in the vessel. By varying the depth, we can measure the velocity profile and thus the flow profile in the vessel. Since ultrasound is propagated at 1500 m/s, processing 1 μ s of signal entails averaging over a depth of 1.6 mm.

The pulse repetition frequency must be chosen with care. All echoes must have been received before the next pulse can be transmitted; deeper lying tissues will generate echoes as well! If pulses are sent with a higher repetition frequency, it will be unknown to which depth a certain echo is related. If the time between pulses is TR, the ultrasound has exactly this time at most to travel from transmitter to reflection site and back. During TR, it travels a distance $TR \cdot c$, where c is the ultrasound velocity in the tissues. Half of this distance is thus the maximum depth.

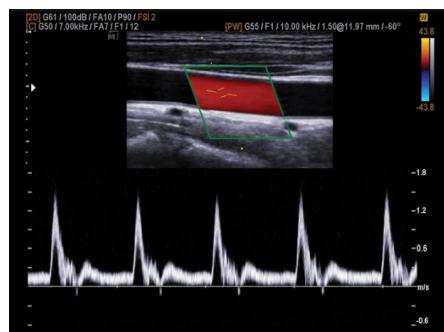


Figure 5-32 Duplex ultrasound image and spectral; Doppler.

In Figure 5-32 an image of the common carotid artery is shown together with a Color Doppler and Spectral Doppler signal. A high-quality image gives information about the best location for the measurement. The measured volume can be selected by the user. The two-dimensional color plot is an indication of the local flow velocity. The spectral Doppler signal of this region is shown below.

5.6 Volume Changes - Plethysmography

Plethysmography is a technique to measure volume changes, it can be used for blood vessels, lung and airways and other parts of the body. The volume changes are often related to changes in blood volume. However, there are other applications such as respiratory measurements and for measurement of body fluids. The three most used techniques are described below.

5.6.1 Cuff based plethysmography

In the 19th century techniques were developed to measure changes in blood volume in extremities *in vivo*. Due to the pulsations of blood in the arteries the volume of arms, legs or fingers change by a small amount. These small volume changes were measured initially with a chamber plethysmograph (see Figure 5-33) that was filled with water. More practical embodiments were developed by Riva-Rocci for measurements of blood pressure at the brachial artery site in the arm. He combined the cuff with a mercury manometer and proposed a criterion for systolic blood pressure, a drawing of an arm cuff is shown in Figure 5-34.

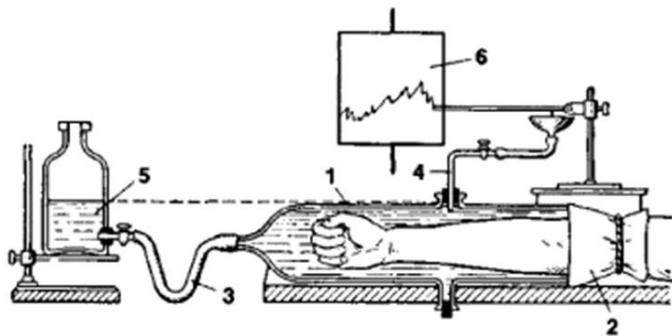
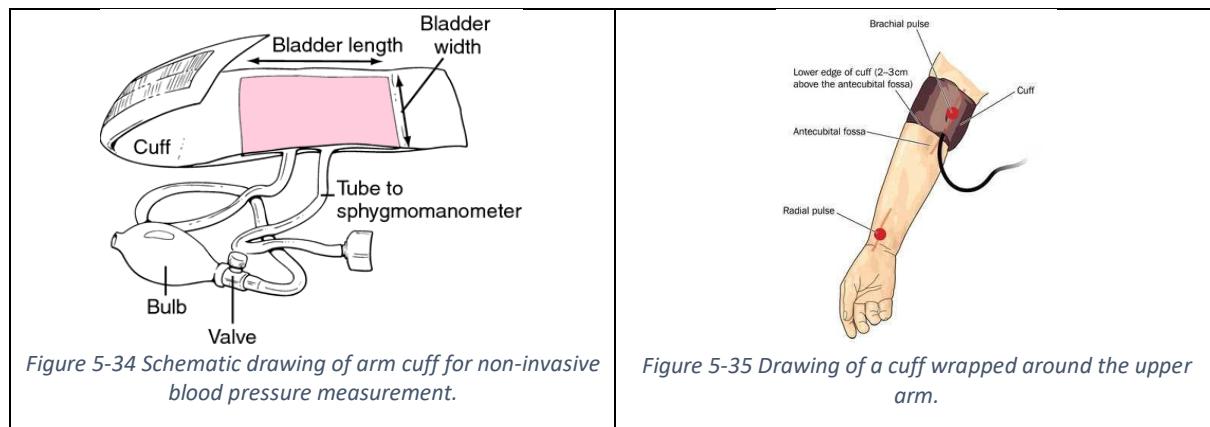


Figure 5-33: Schematic drawing of a chamber plethysmograph to measure small volume changes in the arm.

A cuff consists of an outer layer of a relatively stiff material or fabric with a bladder at the inner side. The bladder width is in the order of 15 cm, the bladder length is about 30 cm. The bladder is in direct contact with the arm. The cuff is wrapped around the upper arm (see Figure 5-35) and the bladder is inflated with air and bladder volume and pressure increase. A typical air volume of the bladder is 200 ml, volume changes of the arm under the cuff are in the order of 0.1ml to 1 ml. The cuff pressure is in the order of 100 mmHg. Volume changes of the arm lead to volume changes in the bladder that lead to small pressure changes (~1 mmHg) in the bladder. These pressure changes are measured and can in principle be converted into volume changes of the arm. The pressure changes are in the order 1

mmHg, about a factor 100 smaller than the bladder pressure. Motion of the arm can lead to much larger pressure changes, the method is prone to motion artifacts. This is a drawback of all plethysmography techniques.



5.6.2 Whole body plethysmograph

For measurement of absolute lung volumes and lung dead space a whole-body plethysmograph is used (see Figure 5-36). This is a rigid and airtight box. The patient sits in the box and breathes air in and out via a mask connected to an external system. During breathing the chest volume changes. This volume change leads to small pressure changes in the box which are measured with great accuracy. These pressure changes can be converted to chamber air volume changes via Boyle's law. The air volume changes in the plethysmograph are equal to changes in thorax and lung volumes. This technique will be discussed in more detail in the chapter of respiratory system measurements.



Figure 5-36 Whole body chamber plethysmograph for measurement of lung volumes, airway resistance and dead space.

5.6.3 Photo Plethysmography

When a light source emits light towards the body, a fraction of the incident light can be transmitted through tissue, the remainder is partially absorbed and scattered by tissues, bones and blood vessels. Arterial pressure pulsations lead to small volume changes of blood in these vessels. Changes of volume of the vessels changes light absorption, reflection and scattering of light. The pulsatile variations in blood volume show up in the form of small light intensity changes of the order of 1 percent of the total detected signal in transmitted or reflective light. This small signal is called the

photoplethysmography signal or in short ppg or pleth signal. The measurements of the volume changes are useful to determine pulse rate and can also be used to measure the fraction of oxygenated hemoglobin when multiple wavelengths are used. The changes in light are correlated to volume and pressure changes but are a poor measure of real volume changes or arterial pressure.

5.6.4 PPG Transducer and Signals

A schematic drawing of a ppg transducer around a fingertip is shown in Figure 5-37. A body location with a dense network of small arteries, arterioles and capillaries is a preferential site for a ppg measurement. The fingertip is one of the best locations, it combines good light transmission with “large” ppg signals resulting in good signal to noise. Other suitable locations are the ear lobes, nose nostril and forehead. In smart watches the wrist is used as the preferred site.

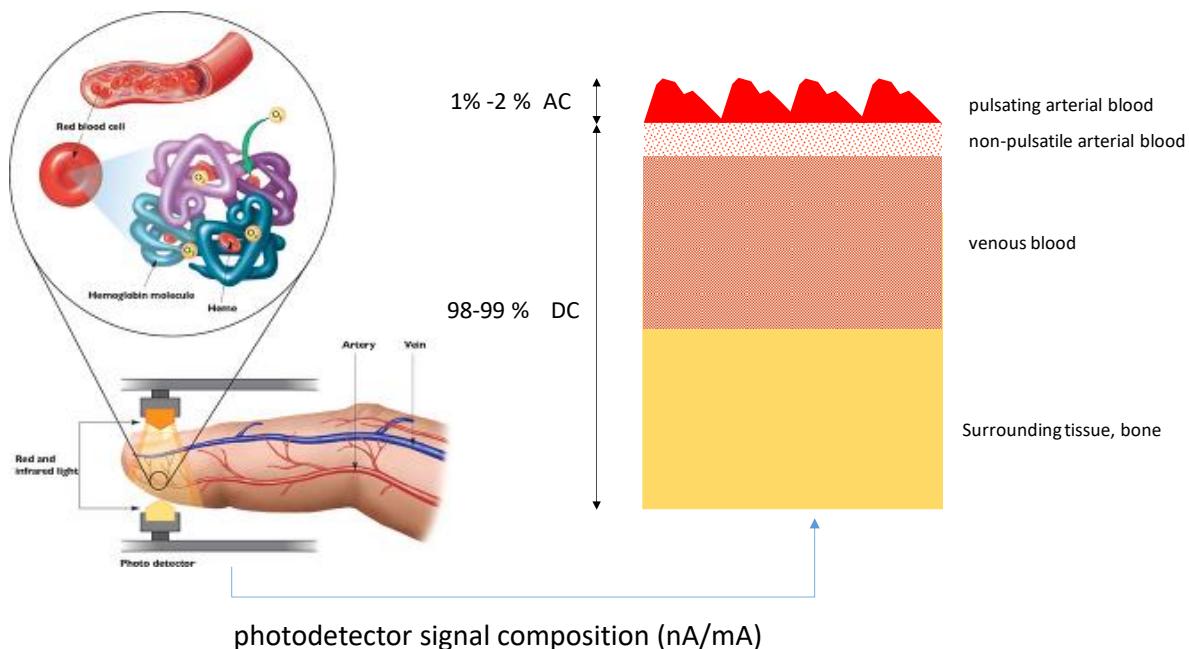


Figure 5-37 Diagram showing the basic features, transducer and distribution of transmitted light.

The sensor consists of a light source and a photo detector. A LED diode is used for the light source, the most used wavelengths are green (530 nm), red (660 nm) and near infra-red (940 nm). The light intensity is a few milliwatt. Green light is absorbed strongly by tissue and blood, it is used mostly for reflective ppg from small arteries near the skin. Red and IR light is used for both transmission and reflective ppg. The light is detected with a silicon photo diode. Optical and electronic noise is a very important issue, both ambient lighting induced noise, LED noise, detector noise and electronics noise need to be minimized. This is a complex task. A simplified schematic of the analog front-end for a ppg measurement or pulse oximetry of a two wavelength system is shown in Figure 5-38.

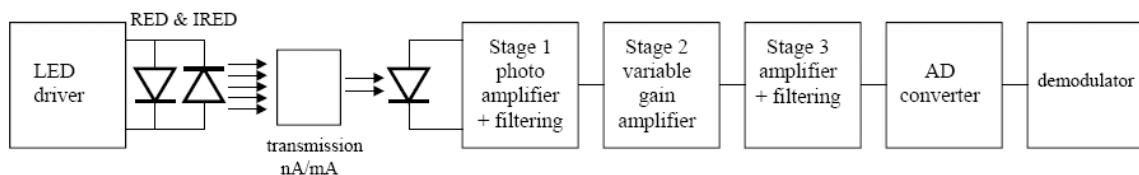
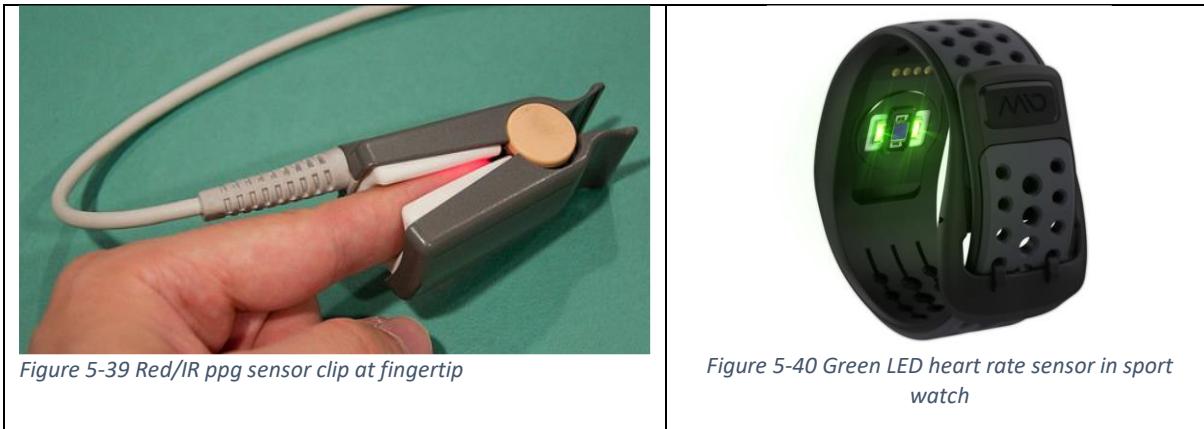


Figure 5-38 Simplified schematic of the electronics of a ppg transducer.

A very low noise pulsed LED driver ($\text{SNR} > 80 \text{ dB}$) is used for the generation of modulated pulsed light from the LED. Modulation has advantages for power consumption, tissue heating, signal to noise and

filtering of ambient light. The transmitted or reflected LED light is detected by a silicon photo diode and photocurrent is amplified and converted to voltage by a low noise trans-impedance amplifier. Further amplification and filtering are done before the signal is converted to the digital domain. In modern devices further processing and feature extraction occurs predominantly in the digital domain. The electronics is preferably adaptive and smart (“digital control”) as ambient conditions and levels of DC and AC amplitudes vary widely.



The red and IR LED devices are mostly used for clinical applications (see Figure 5-39). The green LED is mostly used for heart rate measurements in lifestyle devices such as smart-watches or sport watches. The green ppg measurement location is used most often at the wrist. This location is not optimal but is used for practical reasons. An example is shown in Figure 5-40. Although developed for consumer use such devices may also be useful for home monitoring of people with chronic diseases and sleep disturbances.

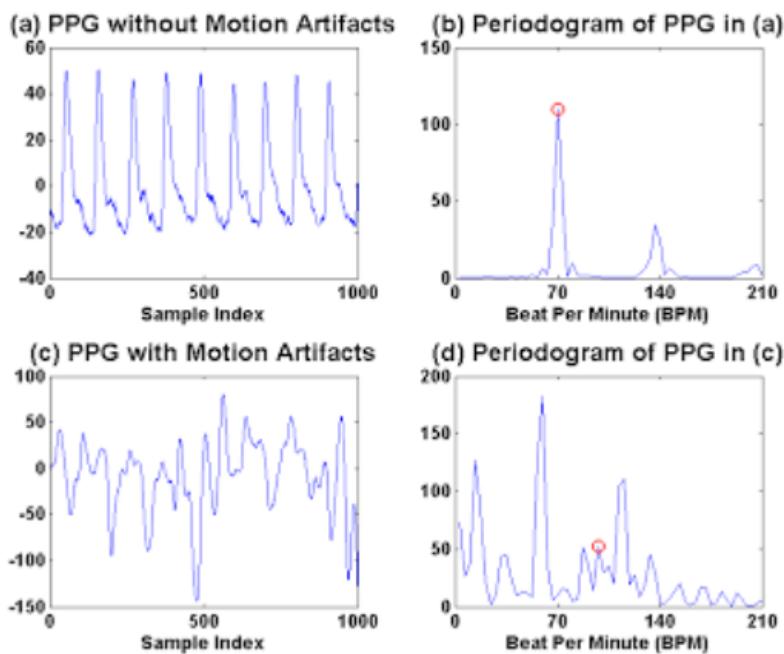


Figure 5-41 PPG signals at rest (a) and with low level motion artifacts (c). The periodograms obtained from Fourier analysis are shown next to the ppg signals

Photo plethysmography is very sensitive to motion artifacts (see Figure 5-41). The pulsatile signal is only a small fraction of the total measured light intensity and motion induced changes in the “DC” part of the transmitted light can easily be ten times larger than that the signal of interest.

Furthermore, the artifact signal can have a very large bandwidth. Note that the motion artifacts shown in Figure 5-41 are relatively small. With advanced signal processing and feature detection techniques heart rates can still be extracted. For severe motion artifacts this is not possible anymore.

The ppg technique is easy to use and gives direct very important clinical information on heart rate, more precisely pulse rate (i.e. a pumping heart, not an ECG). Photo plethysmography was first used in the clinical domain but now also more and more in the lifestyle and sports domains. One of the drawbacks for clinical use is centralization of blood flow when a patient is in a circulatory shock state. Signals at periphery like fingertips and ears are very weak under such conditions and other monitoring sites or techniques must be used.

5.6.5 Electric impedance plethysmography (bio-impedance)

The body consists of tissues and fluids that are relatively good conductors of electrical current (see Figure 5-42). Blood is one of the best electrical conductors in the body and pulsatile changes in blood volume modulate the local resistance of tissue. Furthermore, during breathing the resistivity of the lung and thorax areas changes. The technique is also used for the determination of the percentage body fat (lifestyle, population average but not a reliable personal measure). Resistance changes contain information on blood and air volume changes, and this is exploited in the clinical domain. One of the main applications is the measurement of the respiration rate but cardiac impedance also has been studied extensively.

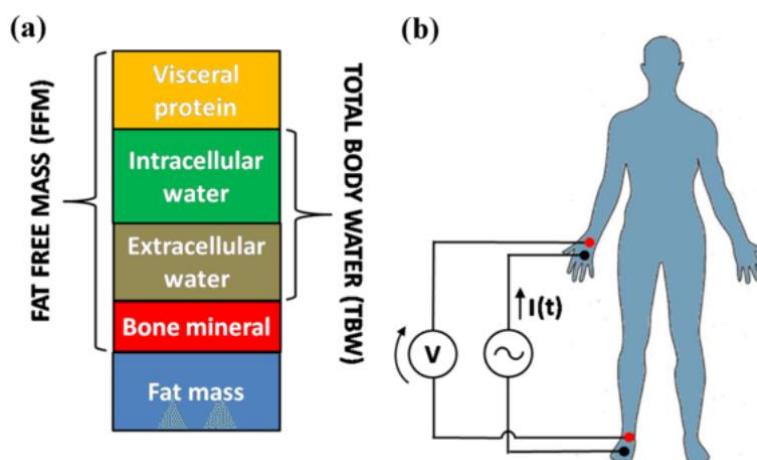


Figure 5-42 Body composition and schematic drawing of bio impedance measurement.

The measurement is straightforward. Electrodes are placed at the skin and small non harmful AC signals (voltages less than 60V, currents < 1 mA) are injected into the body and the resulting voltages are measured. The bio-electrodes differ from metal contacts, in the body the charge is carried by ions rather than by electrons (see Figure 5-43). Therefore, special bio-electrodes have been designed that convert ion current into electron current. The most used electrode is a silver-silver chloride gel electrode. These contacts are in fact electro-chemical half cells. As a result, there are half-cell DC offset potentials in the order of a few hundred millivolt. The contact resistance depends strongly on frequency (see Figure 5-43) and is sensitive to motion artifacts (slow ion motion in skin layer). At DC frequencies the contact impedance is of the order of tens of $k\Omega$. Above 30 kHz a low value of 100 Ω is reached and contact resistance is almost constant. Because of these properties most bio-

impedance measurements are done in the frequency range 30 kHz to 50 kHz. Higher frequency poses other problems. Often the changes in resistance are so small that even a contact resistance of a few hundred Ohm poses problems. This can be solved by using 4-wire impedance measurements. Two contacts are used for current injection, two separate contacts are used for the voltage measurement. By using a very high impedance voltmeter, currents flowing into the voltmeter can be made negligible small and potential drop over the electrodes can be neglected.

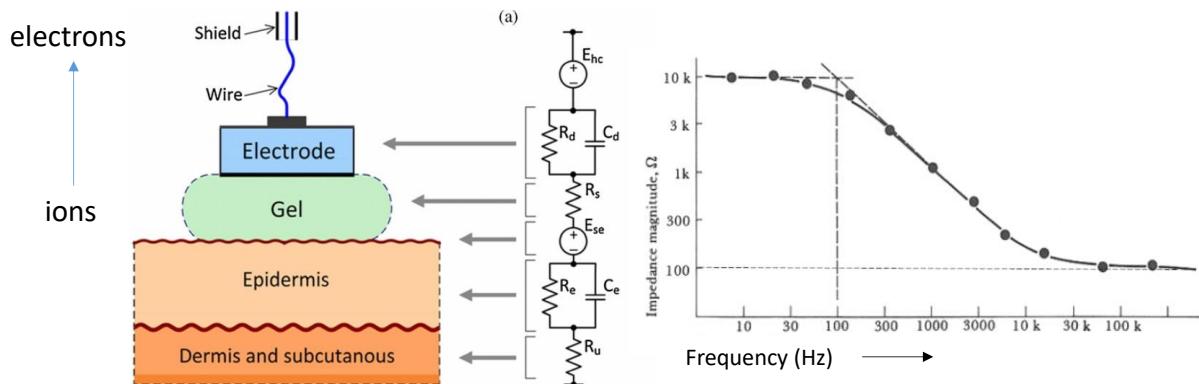
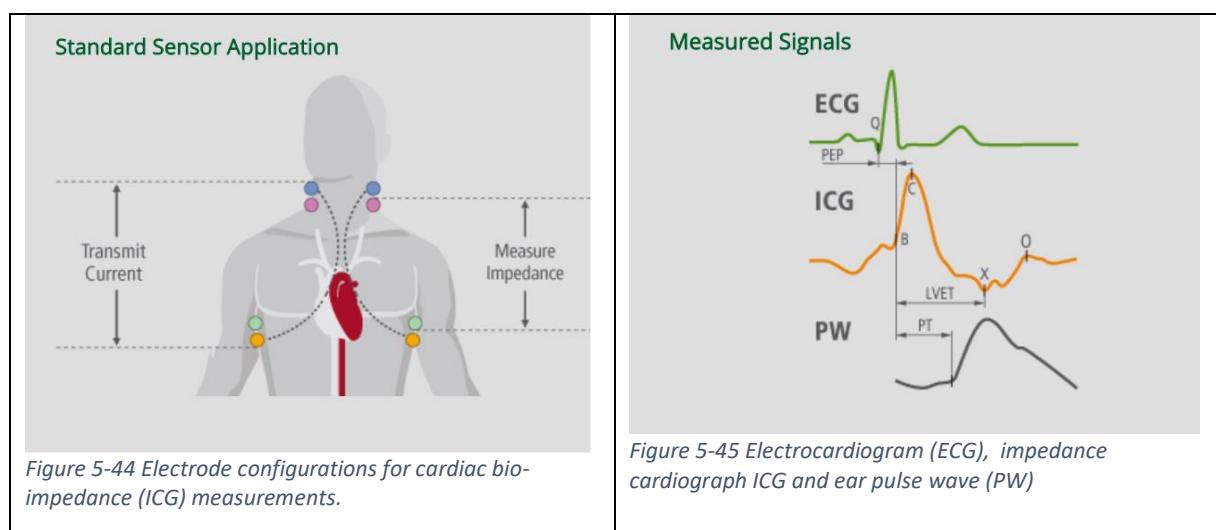


Figure 5-43 Diagram showing the skin composition, ECG Ag/AgCl electrode, lumped element model of the contact and skin and contact resistance versus frequency of an Ag/AgCl electrode.

Typical electrode arrangements for impedance cardiography (ICG) are shown in Figure 5-44. Eight large electrodes are placed at optimal locations at the neck and side of the thorax just below the heart. Four are used for AC current injection and four are used for 4-wire impedance measurements. From the non-invasive ICG signal important parameters such as left-ventricular ejection time (LVET), pre-ejection period (PEP) and pulse transit time (PT) can be extracted. Note that there is still controversy on the accuracy of these parameters when determined from ICG signals.



In most monitors bio-impedance is used for the extraction of respiration rate during an ECG measurement. The same electrodes as used for the ECG measurement can be used for the impedance measurement. The ECG and impedance signals are in the low (< 1 kHz) and high frequency bands (> 30 kHz), signals can be separated easily by filtering. The bio-impedance measurement is a two-wire measurement, therefore contact resistance is included in the measurement. A bio-impedance signal is shown in Figure 5-46 for a patient who has a period of apnea. The capnography signal shows when the patient breaths. During inspiration a 0.6 Ω resistance

increase is observed. The DC resistance was approximately 150Ω . During the apnea period a small signal of 0.1Ω is observed, it is caused by the heart beats. At the end of the apneic period the patient moves and large artifacts on the impedance signal are observed. Bio-impedance is extremely sensitive to motion artifacts. Note that after movement the DC level has changed.

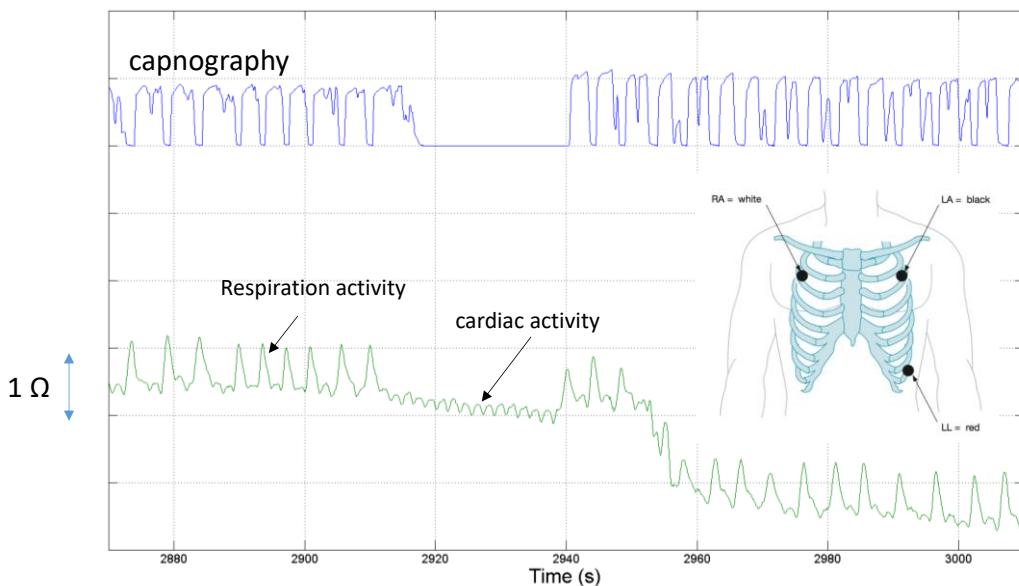


Figure 5-46 Chest bio-impedance and capnography data of a ventilated patient.

5.7 Electrocardiogram

As was described in the physiology lectures there is an electro-mechanical coupling in the heart that causes the mechanical contraction of the heart. The polarization and depolarization of the heart muscle cells produce small time dependent electrical currents in the thorax. These currents produce small body-surface potentials (peak value $\sim 1\text{mV}$) which can be measured during the heart cycle [Guyton]. Electrocardiography (ECG) is the recording of electrical activity of the heart over time by measuring small potential differences between electrodes placed over the chest. The locations are not arbitrary and there are standard locations [Guyton]. The measured potentials and signals depend strongly on the location of the electrodes both in shape and magnitude. For clinical diagnosis standard locations of the electrodes are mandatory. ECG measurements are an important clinical measurement for diagnosis of heart pathologies. For diagnosis, the fidelity of the measured signal is extremely important. Both its wave shape and its magnitudes are of great importance. Therefore, the demands on the ECG measurement are severe and tight specifications are described in international standards. A not complete list of the requirements can be found in [Webster].

Einthoven was a Dutch physician who invented the string galvanometer and was able to measure the small ECG potentials. He was the first to study the ECG in detail and is the inventor of the ECG measurement and use in clinical diagnosis. He measured potential differences between the right arm, left arm and left leg. The potential difference between two contacts is called a lead. With three electrodes in the Einthoven arrangement three leads can be measured, they are called lead I, lead II and lead III. This is the Einthoven triangle. An ECG lead measures the component of the current in a certain direction as indicated in Figure 5-47. The ECG signal of a lead is representative of the electrical activity of specific parts of the heart. By measuring different ECG-leads a mapping of the electrical activity of the heart can be obtained. Local electrical abnormalities of the heart muscle show up as

specific features in the ECG signals. By measuring multiple leads the location of this region can be determined.

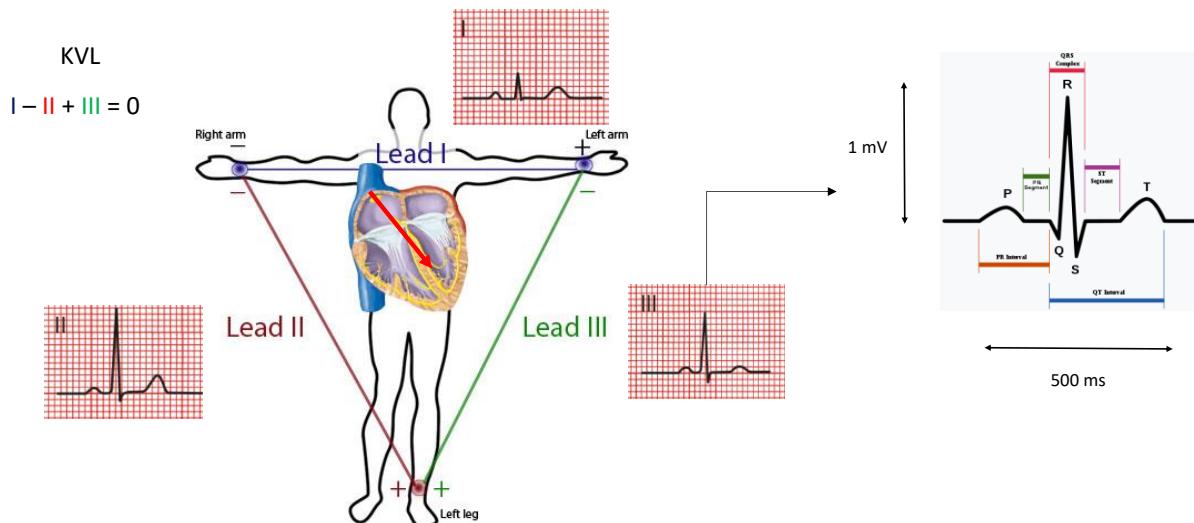


Figure 5-47 Einthoven's triangle and corresponding ECG signals. A schematic drawing of an ECG signal with nomenclature is shown

A schematic drawing of a lead III ECG trace is shown in Figure 5-47. The P-top is a small (~ 0.2 mV) signal that is representative of the electrical activity of the atrium, the QRS complex is representative of the depolarization of the ventricles, it is the largest signal in the ECG (~ 1 mV). The T-wave is representative of the repolarization of the heart. As note before both the amplitude and the faithful reproduction of the shape of the ECG signal are of importance for the clinician. This leads to demands for the distortion and bandwidth of the ECG measurement. The sharp PQRS complex requires a measurement bandwidth in excess of 150 Hz.

The small ECG signal (1 mV) is measured in a very complicated and noisy environment. The electromagnetic noise level (mainly 50 Hz – 60 Hz power lines) can be very high. The diagram illustrates the potential issues that needs to be tackled for a high-quality clinical ECG measurement. The ECG electrodes are connected via shielded cables and via a trunk (contains high power protection resistors of $1\text{ k}\Omega$) with one shielded multi-stranded cable to the patient monitor. Capacitive and magnetic coupling with powerlines can induce large 50Hz common mode signals at the input terminals of the ECG measurement electronics. The skin contact impedance of the ECG electrodes is large ($\sim 50\text{ k}\Omega$ for low frequency signals). The contact impedance Z can vary from electrode to electrode. Z also depends on the quality of the placement procedure and varies with time of use. Besides the ECG signal additional voltages are present, a not-complete list is: motion sensitive DC electrode offset voltages (variable ~ 200 - 300 mV), higher frequency electro-myography signals due to muscle activity, powerline interference (50Hz, 60 Hz) and electro-magnetic interference from analog and digital devices (laptops, tablets, monitors, surgical instruments) and mobile devices. The magnitude of these noise signals depends on many factors, differences in electrode contact resistance is an important one. A careful design and selection of components can reduce this interference but not eliminate them. Note that noise signals can be orders of magnitude larger than the ECG signal. When a pacemaker is present, the high frequency heart trigger pulses produced by the pacemaker need to be detected and must be marked at the monitor display and must be removed from the ECG trace.

The patient monitor is connected to mains supply and is grounded to the hospital ground connection. The DC supply of the measurement must be electrically isolated from mains. Powerline noise from

the mains supply must be very small. At single fault conditions leakage currents flowing through the patient must be smaller than $10 \mu\text{A}$ as currents as small as 10's of μA can be lethal under certain conditions.

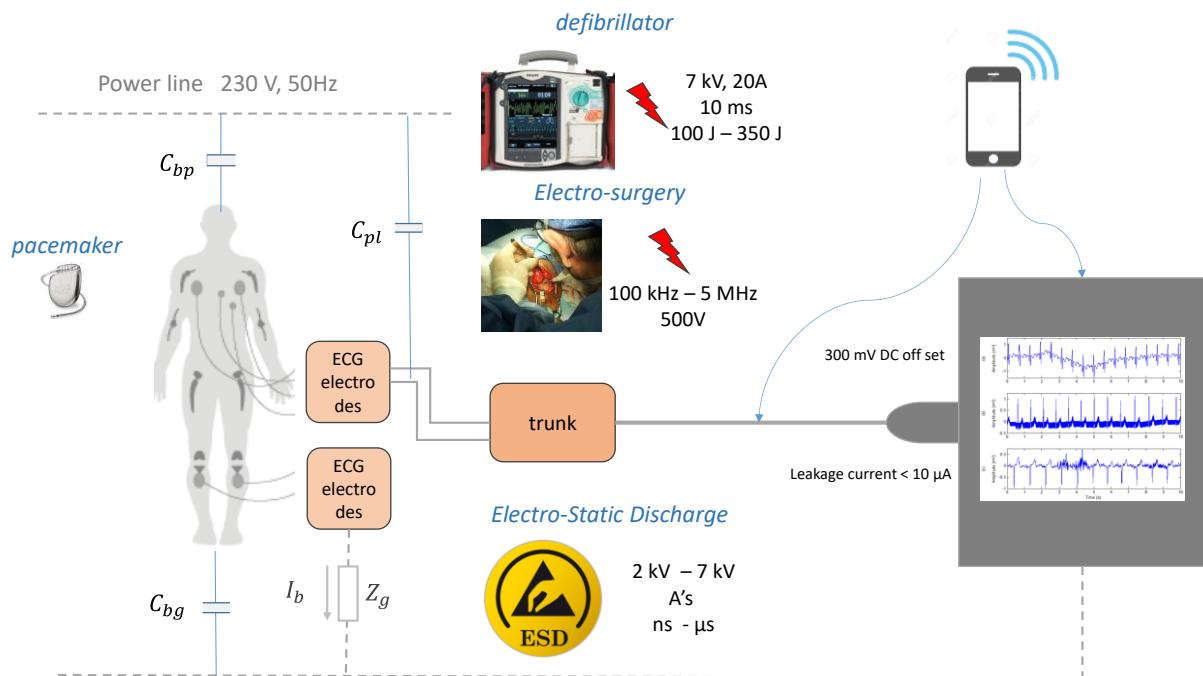


Figure 5-48 Diagram showing the environmental disturbances and safety aspects. Different disturbances to the ECG signal are shown in the insert (base line wander, 50 Hz pick-up and muscle activity)

During clinical care additional and potentially dangerous electro-magnetic interference can occur. When a defibrillator is used a short high-voltage high-current pulse flows through the thorax of the patient. The energy in the pulse varies between 100J and 350 J. The pulse duration is in the millisecond range. Voltages can be as high as 7000V in old systems. Part of this high-voltage high-power pulse partly can flow through the ECG measurement electronics and can permanently damage the electronics. The electronics must be protected from this defibrillation pulses. This requires several measures which will be discussed in a following section. A second hazard occurs when a surgeon uses electro surgery knives to make incisions in skin and tissues. These tools use high power RF to make cuts, typically frequencies are in the range of 100 kHz to 5 MHz and voltages can be in excess of 500V. These voltages can easily damage the electronics and can also disturb and suppress the ECG measurement. Finally, static electricity can charge objects and persons and high voltage ESD discharge pulses can occur. Dangers for the patient and electronics need to be dealt with.

5.7.1 Input protection

The measures taken to protect the measurement electronics against high voltage pulses are shown schematically in Figure 5-49. The first protection layer is the trunk resistor which limit the current that can flow towards the measurement. The resistance is about $1 \text{ k}\Omega$, the resistances are able to withstand the high voltage and high-power dissipation during a defibrillation shock. These are large and expensive components. The resistance value is chosen as a compromise between power loss from the defibrillation pulse and compatibility with bio-impedance respiration rate measurement (DC series resistance). The connector must be able to withstand high voltages and have low leakage and creep currents between the pins, therefore the connector is large and bulky and expensive. The next stage is the ECG amplifier protection from the surge current (amperes) and high-voltage (1000's volts) from the defibrillation pulse. This stage consists of either gas discharge lamps or sidactor

protection elements (type of a thyristor). This stage diverts the high voltage/current pulse towards the ground lead and from there back to the patient. A large fraction of this current will flow back to the patient. The following stage protects the electronics from short duration high voltage (kV's, nanoseconds) electro-static-discharge pulses that can be induced by static charging of objects and persons. The next stage consists of two low-pass filters that further limit inrush currents and that also function as a second order low pass filter for the first amplifier stage. The resistance values are in the order of 10's of $k\Omega$. The resistances also limit leakage currents that can flow from the electronics to the patient.

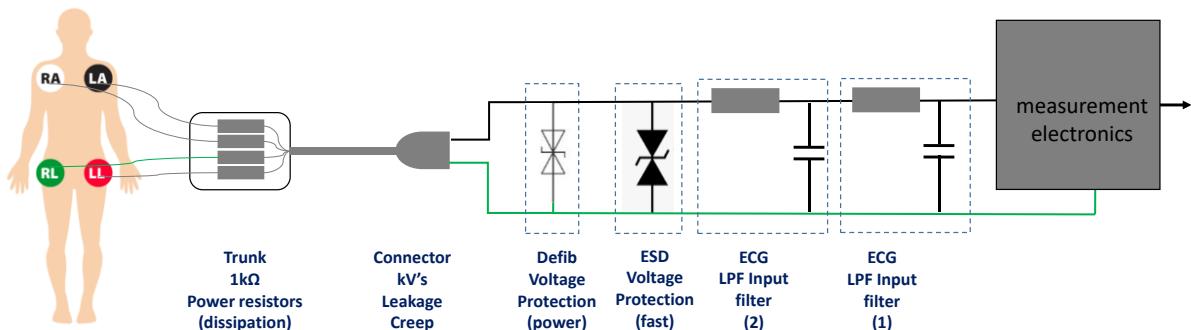


Figure 5-49 Diagram of patient and electronics voltage protection elements.

5.7.2 ECG Amplifiers

The classical approach for a differential ECG amplifier is shown in Figure 5-50. The lead selector and protection devices are not included. The total gain must be in the order of 1000 ($1mV \rightarrow 1V$).

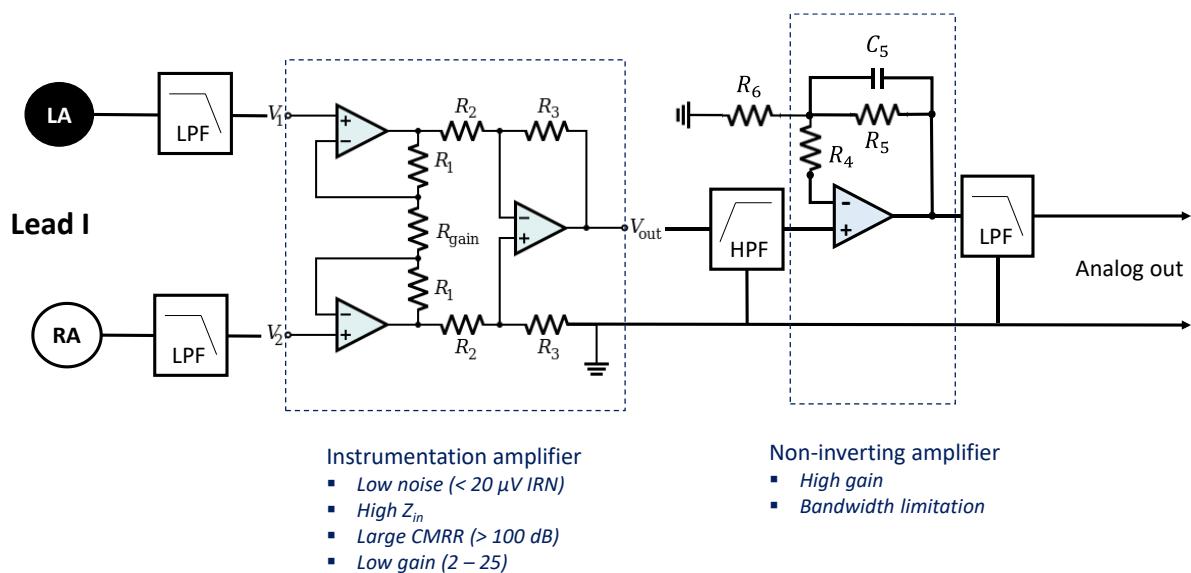


Figure 5-50 Differential ECG amplifier with low gain instrumentation amplifier as input stage and an AC coupled high gain second stage amplifier.

The first stage of the amplifier or preamplifier is a low-gain high common mode rejection differential amplifier ($G_p < 25$). It is coupled via the low pass filter to the ECG electrodes to remove the high frequency noise. The first stage must have very low noise levels (this stage determines signal-to-noise) and high common-mode voltage rejection (CMRR, DC and 50Hz). The input referred noise

(IRN) must be smaller than $20 \mu\text{V}$. Furthermore, a high input impedance at low frequencies is required ($> 2.5 \text{ M}\Omega$ @ 10 Hz) to reduce impact of noise currents. The gain of the preamplifier must be low to eliminate saturation effects, this stage also amplifies the large DC offset voltages and other low frequency (50Hz) noise voltages. The input of the second stage of ECG amplifier is AC coupled to the output of the first stage (high pass filter, 0.05 Hz cut-off frequency) to remove the DC offset voltage and guarantee a high level of electrical isolation from the front stage and the supply voltage unit. After bandpass filtering a high gain is possible to obtain an output voltage in the order of 1 to 10 V . The bandwidth is determined by the RC time of the output stage (set by R_5 and C_5). The upper cut-off frequency should be larger than 150 Hz . The analog output can be connected via isolation devices and buffer amplifiers to an output recorder. Alternatively, the output signal can be digitized and processed further in a computer system or micro controller. The disadvantage of this ECG amplifier are its high power consumption, high cost and large size and a high count of expensive parts.

A modern implementation of a clinical ECG amplifier is shown in Figure 5-51. It is not a complete schematic; some parts are not shown (Wilson terminal etc.) The operating voltage can be low (2.5V or 3.3V). The ECG electrodes are connected via the low-pass filter to the non-inverting input of a buffer amplifier. This amplifier must have a very low-gain ($G < 5$) to avoid output saturation, must have a very high-input impedance, common mode suppression and must use very low noise components. The outputs of these buffer amplifiers are connected to a multiplexer and are connected via an anti-aliasing filter to a differential ADC converter (16-bit – 24-bit, sampling frequency $> 15 \text{ kHz}$). The different channels can be sampled in succession. The sampling frequency per channel is a few kHz . Further filtering, signal processing, sample rate conversion and artifact removal is done in the microcontroller. The ECG lead signals on the display can be obtained from the digital signals. The advantages of this architecture are a low part count, small area, relative low power consumption¹⁰ and low cost. By using this more digital implantation of an ECG amplifier more flexibility is possible and functionality per use case can be changed by software.

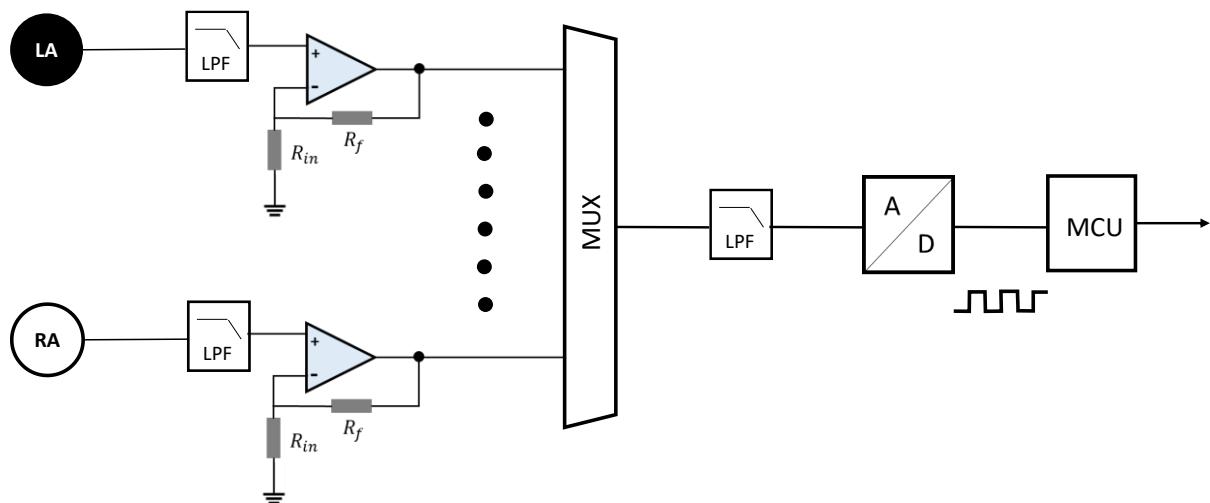


Figure 5-51 Digital ECG amplifier architecture.

¹⁰ Due to modern high performance IC's (ADC, MCU)

5.7.3 Additional ECG circuit Blocks

A high-level diagram of additional blocks that are present in state-of-the-art ECG measurements are indicated in Figure 5-52.

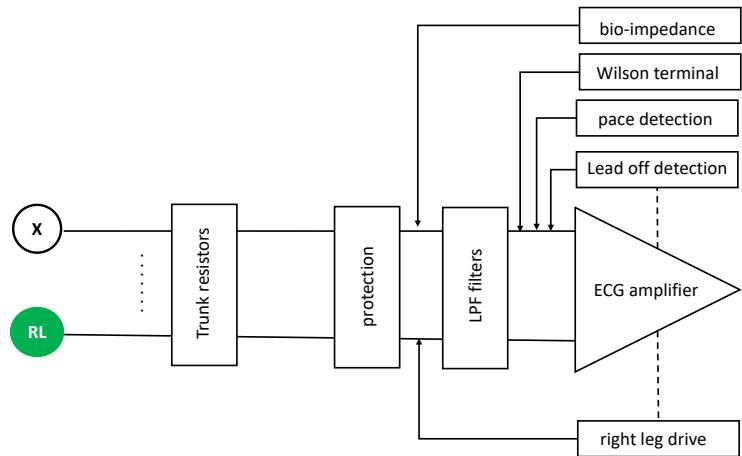


Figure 5-52 High level diagram of other circuit blocks included in a modern ECG measurement.

The patient is not grounded in a modern ECG system. The right leg drive (RLD) output voltage is connected to the right leg electrode. The RLD feedback voltage ensures a clean earth contact to the patient. The measured common mode voltage is inverted, amplified and fed back to the right-leg contact. The negative feedback reduces the common mode voltage to a small value. The 50Hz capacitive coupling current now flows to the ECG unit.

Most ECG systems verify if an electrode is connected (lead-off, INOP condition). In some devices the quality of the contact is measured. Warnings and alarms are given when a lead-off condition or a high contact resistance is detected.

For several ECG leads a reference voltage is required. This is called the Wilson terminal. It is a sort average of all electrode potentials; it is obtained by adding currents of different electrodes and further processing.

In a clinical grade ECG unit pulses from a pacemaker must be detected, removed from the signal and indicated in the ECG trace on the monitor display. The pace pulses vary per pacemaker type and per manufacturer and are often short duration pulses that require higher bandwidth for detection than is needed for the ECG signal. Therefore, often a special circuit block is included for the pace pulse detection.

Most ECG devices include a bio-impedance measurement to measure the respiration rate. The same ECG contacts are used for this measurement. The sensing voltage is applied after the low pass filters, the additional series resistance of these filters is too large for the measurement of the respiration induced thorax bio-impedance.

5.8 Temperature

The temperature of the central part of the body is an important clinical parameter. Body temperature is tightly regulated and varies somewhat over the day. Normally it is in the range between 36°C and 37.5°C. The required accuracy of the measurement is 0.1°C. It should be noted that body temperature is not a well-defined parameter, there is a wide range in temperatures inside the body and on the skin. Some are of high clinical interest (the core temperature) and others are irrelevant (skin temperature). This is illustrated in Figure 5-53 (a).

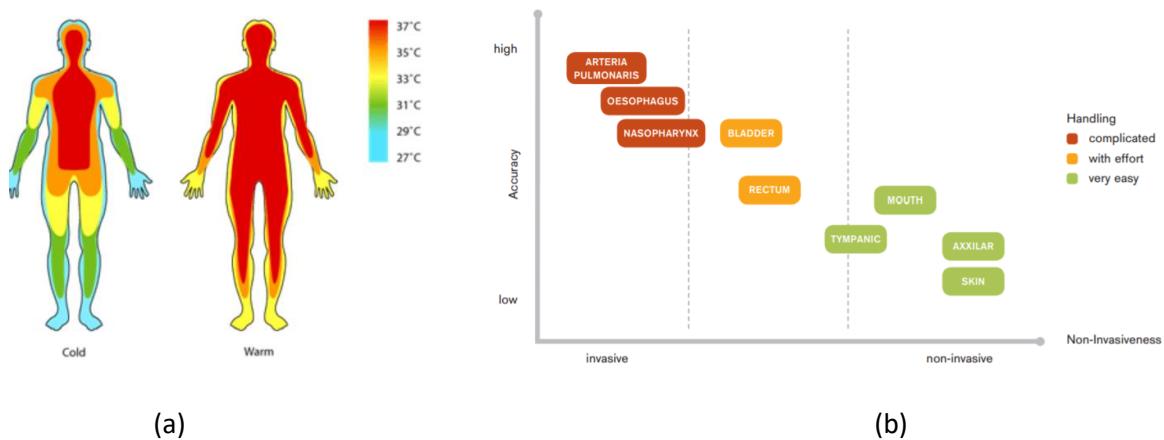


Figure 5-53 Diagram showing the temperature distribution in the body (a) and possible locations marked with accuracy and degree of invasiveness (b).

The lowest and highest values can differ by several degrees. The temperatures at the extremities and the skin are lowest. The temperature in the core of the body (brain, thorax, and abdomen) is highest. Differences can be as large as 10 °C. Clinicians want to know the core temperature; they are not interested in skin temperature because this is normally much lower than the core temperature. Unfortunately, the core temperature is difficult to measure and requires more or less invasive methods (see Figure 5-59 (b)). The invasive locations are reached by sensors placed in catheters.

Historically the mercury thermometer was the device of choice for temperature measurement. It has been replaced by electronic sensors for safety and workflow reasons. Thermistors are the most used temperature sensor for high accuracy clinical applications (Figure 5-54 (a)). A thermistor is an electrical resistor whose resistance varies with temperature. A table or polynomial fit is used to convert resistance into temperature. There are thermistors with a positive (PTC) and negative temperature coefficient (NTC) of resistance (Figure 5-54 (b)). For a PTC thermistor the resistance rises with temperature, for a NTC resistor resistance decreases with temperature. Most characteristics are non-linear. The metal resistors (e.g. platinum resistor) have a nearly linear transfer characteristic. Materials used are semiconductors, metal oxides, ceramics and metals. Thermistors are available in a wide variety of sizes from large high-power devices to miniaturized components (Figure 5-54 (a)). The symbols for PTC and NTC thermistors are shown in Figure 5-54 (c).

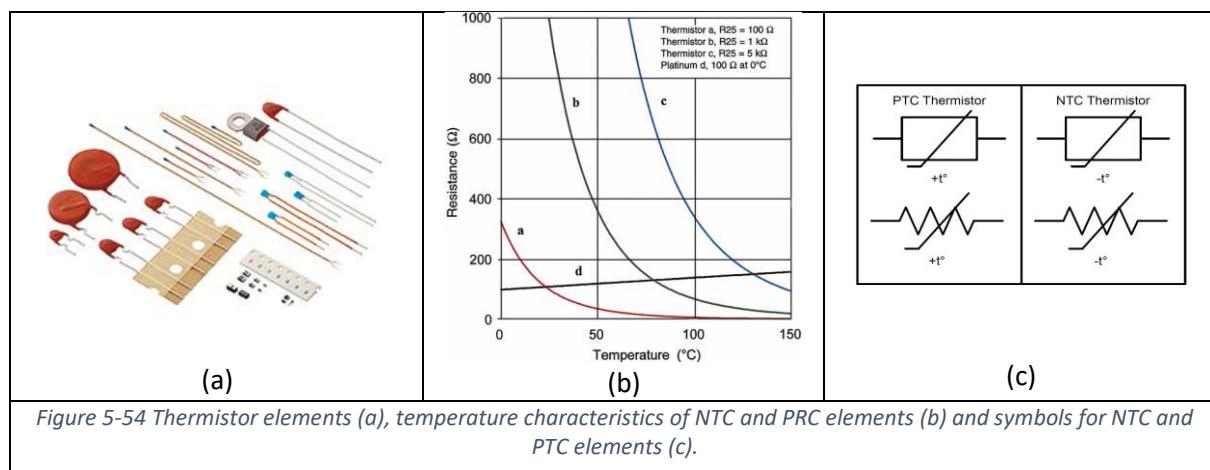


Figure 5-54 Thermistor elements (a), temperature characteristics of NTC and PRC elements (b) and symbols for NTC and PTC elements (c).

The measurement of the resistance is done in the Ohmic regime, i.e. the measured voltage and current should be low enough to avoid non-linearity due to heating. A simple low-cost implementation that is used in some patient monitors is shown in Figure 5-55.

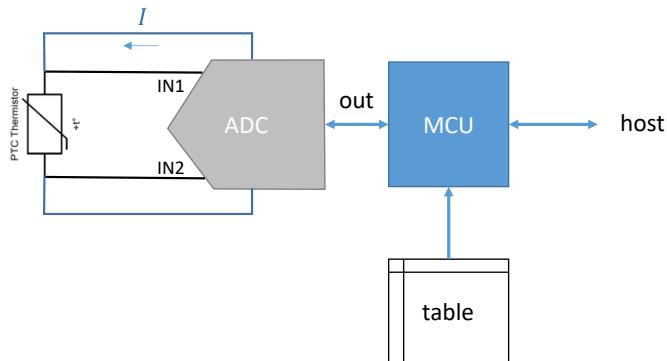
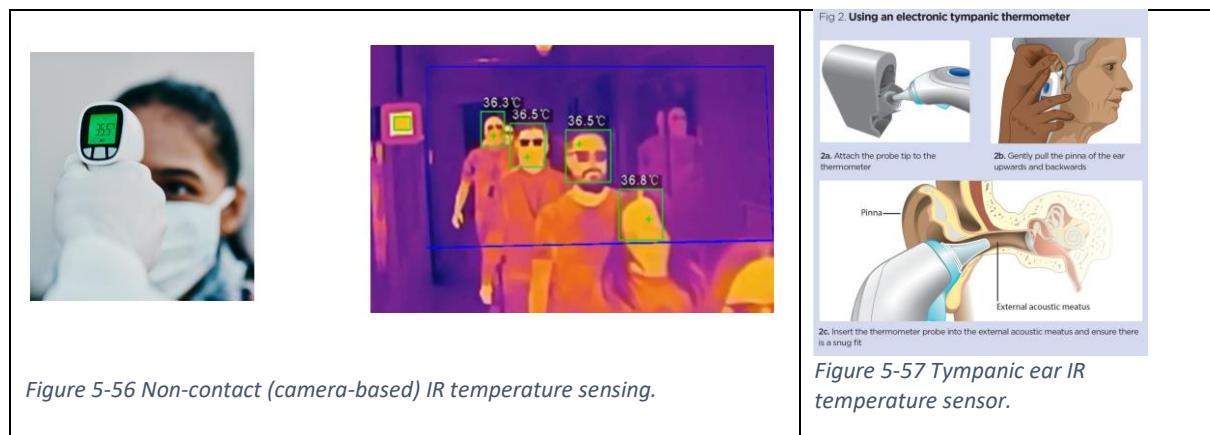


Figure 5-55 Diagram of a modern implementation of a "digital" temperature measurement.

Modern analog-to-digital converters (ADC) have multiple hardware blocks, a current source is often included. The ADC sends a constant current through the thermistor and the voltage over the thermistor is fed into the differential input of the ADC. The output of the ADC is fed into a microcontroller that converts the measured voltage into a temperature reading. The MCU is connected to a host device (e.g. a patient monitor). This measurement is simple, flexible and low-cost and is very accurate.



A second sensing method that is widely employed is the infra-red (IR) temperature measurement. An IR temperature sensor extracts the temperature from a specific part of the body and measures part of the thermal energy of black-body radiation emitted by an object ($5 \mu\text{m} - 12 \mu\text{m}$ wavelength). It is a non-contact sensor and the temperature reading is available in seconds and is shown on a digital display (see Figure 5-56). The device is calibrated for a fixed emissivity (range 0.97 to 0.99). The accuracy and precision are less good than that of the thermistor probe. For skin temperature the required accuracy is $\pm 0.3^\circ\text{C}$ and medical quality devices must also include a so-called adjusted mode where the core temperature is estimated. The difference between skin and core temperature at the forehead can be as large as 3°C . A tympanic IR ear thermometer is most frequently used in low acuity parts of hospitals like the ward. The temperature of the ear drum is probed. This location is closer to the brain and might be a better approximation of the core temperature. This device is very convenient to use and accuracy is often good enough for screening purposes and spot check measurements (when the IR ear thermometer devices are used). Note however that errors can be

very large when emissivity of the object differs from the reference value, when there is ear smear, when distance is too large or when the location is not well defined.

Three embodiments of temperature measurement probes that are widely used in clinical practice are shown in Figure 5-58. There are probes for skin and rectal temperature (a), an IR sensor for tympanic temperature (b), this location in the ear is closer to the brain and invasive sensors for measurement of a specific core temperature (c). Note that incorrect position in the ear and ear smear can increase the measurement error ($> 0.5 \text{ C}$) The rectal probe does not measure core temperature and has a lag in time with the actual temperature. The bladder probe is not patient friendly and is used during treatment when a bladder catheter is required for other purposes. The esophageal probes are mostly used during anesthesia. The invasive sensing solutions for core temperature cannot be used during low level care.

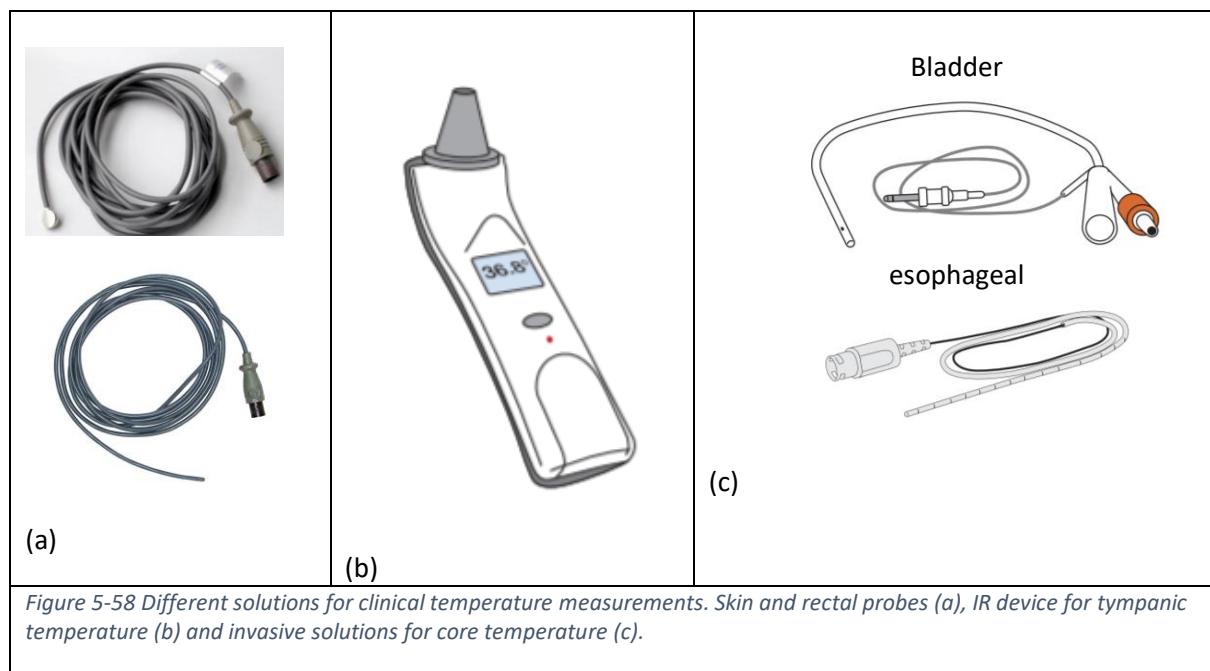
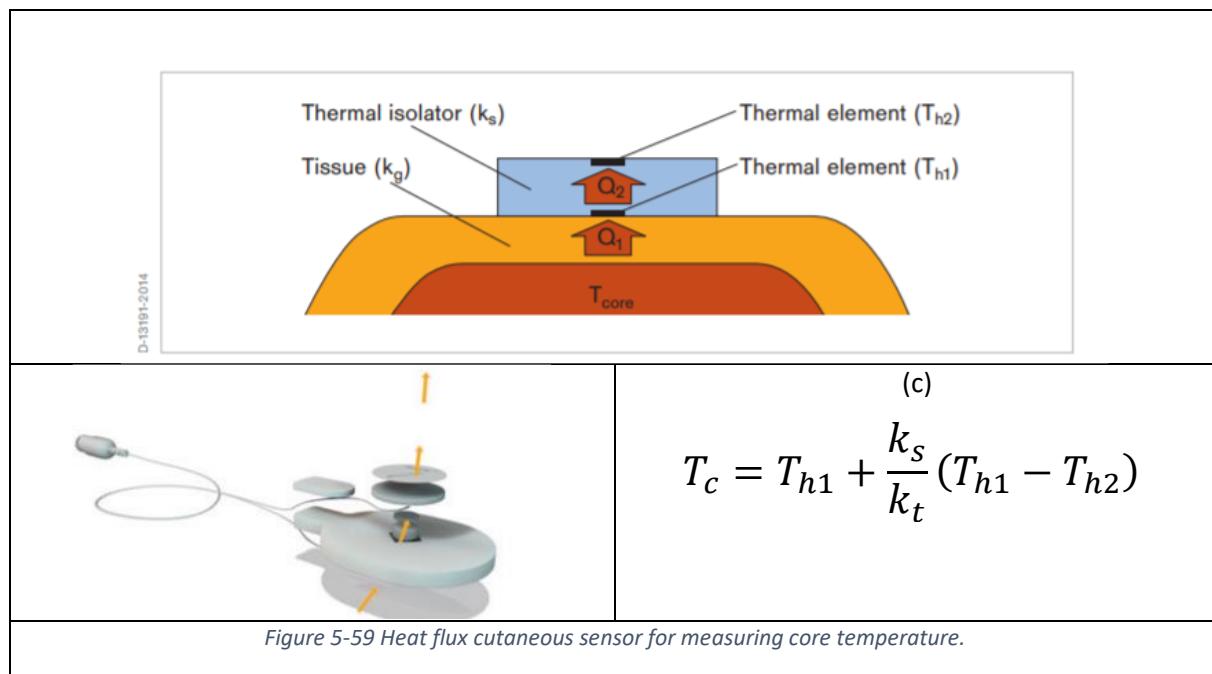


Figure 5-58 Different solutions for clinical temperature measurements. Skin and rectal probes (a), IR device for tympanic temperature (b) and invasive solutions for core temperature (c).

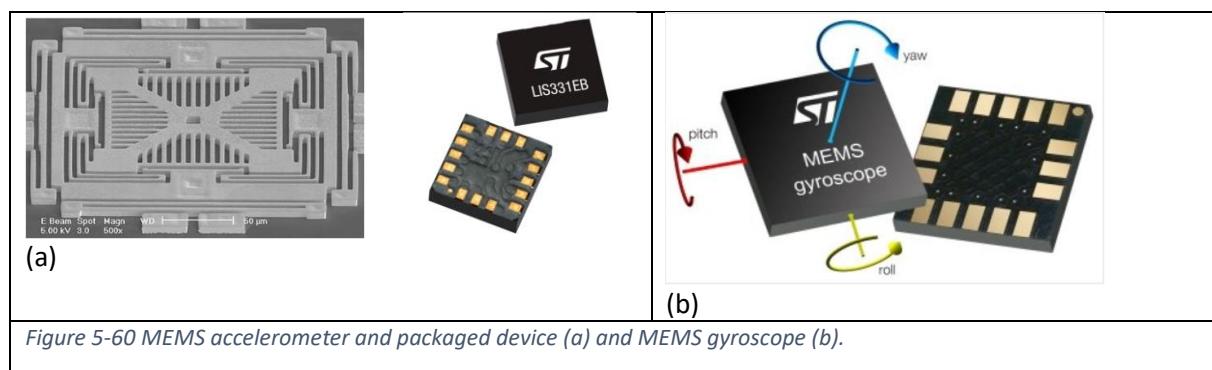
There is a need for a non-invasive brain core-temperature sensor. The zero-heat flux sensor has been studied in the past. The measurement principle is that skin and core temperature are the same when there is no heat flowing from the brain to the skin. This can be measured by active probes using carefully designed devices with multiple temperature probes and a high degree of thermal isolation from the environment. There have been many attempts but recently non-invasive cutaneous sensors for the core-brain temperature have become commercially available. The principle of operation is illustrated in Figure 5-59. The sensor is a variant of the zero-heat flux sensor. When two objects have the same temperature the net heat flow between the objects is zero. In this device there is no temperature regulation but the measured temperatures and the known thermal properties enable an estimate of the core temperature (see Figure 5-59 (c)). These devices are not widely used yet, they are now tested on a larger scale and there are still issues with the accuracy and precision.



5.9 Motion and Acceleration

MEMS accelerometers and gyroscopes are widely used in cars (crash detection), navigation systems and smart phones. They are now included in almost every smart phone, tablet and smart watch. The performance has increased while cost per unit has dropped to very low levels (< 1 €). They have also found use in the medical domain. The main features and performance are described below.

An accelerometer is a device that measures the acceleration of the unit in its own coordinate frame. In rest an accelerometer measures an acceleration in two or three directions due to the earth gravity ($9.81 \text{ m}^2/\text{s}$).



The acceleration is measured with a capacitive technique. Most devices measure the acceleration in two or three directions (3-axis accelerometers are most common). The dimensions of packaged devices are small, a typical volume is a few cubic millimeters (Figure 5-61). The electronics and ADC convertors are integrated with the sensor and are in the same package, the output signal is most often digital. The supply and variety in device properties is huge. Power consumption is very low, a few microwatts is possible (Figure 5-61). MEMS accelerometers with a 2g range and 16-bit resolution, bandwidth above 100 Hz and power consumption of 10's of microwatt are available and are used in medical applications. This will be described in the lecture on low acuity monitoring.

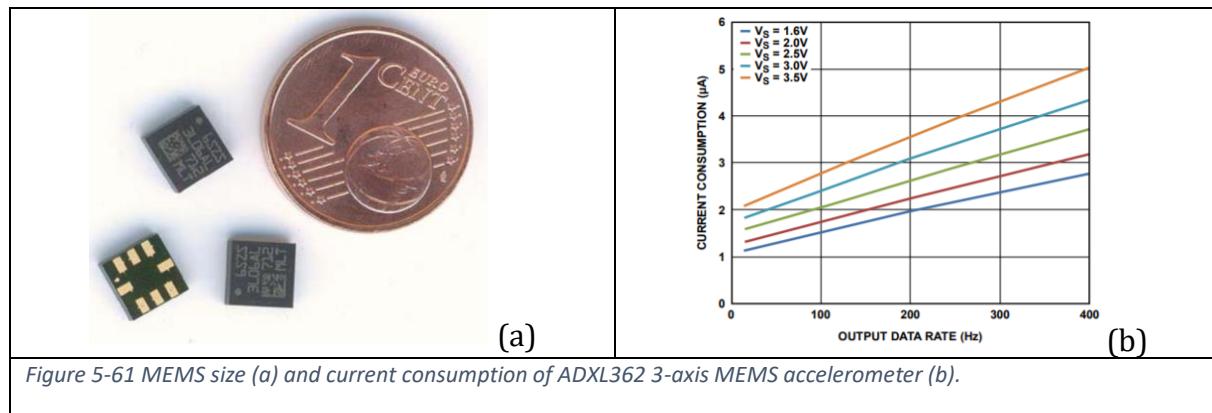


Figure 5-61 MEMS size (a) and current consumption of ADXL362 3-axis MEMS accelerometer (b).

5.10 Capnography

Capnography is the monitoring of the partial pressure of CO_2 in inspired and expired air close to the airway opening (a mask, endotracheal tube or nasal cannula is required). A typical concentration of CO_2 in expired air is around a few percent (i.e. 30 – 40 mmHg). A device that measures the CO_2 pressure during breathing is called a capnograph. There are several sensing methods to measure CO_2 pressures including electrochemical and optical techniques. Demands on the accuracy, precision and speed of the measurement (rise time ~50 milliseconds) rule out electro-chemical sensing techniques and optical techniques are the preferred choice for capnography. The most used optical technique is non-dispersive infra-red (NDIR) absorption. A schematic drawing of a NDIR system is shown in Figure 5-62.

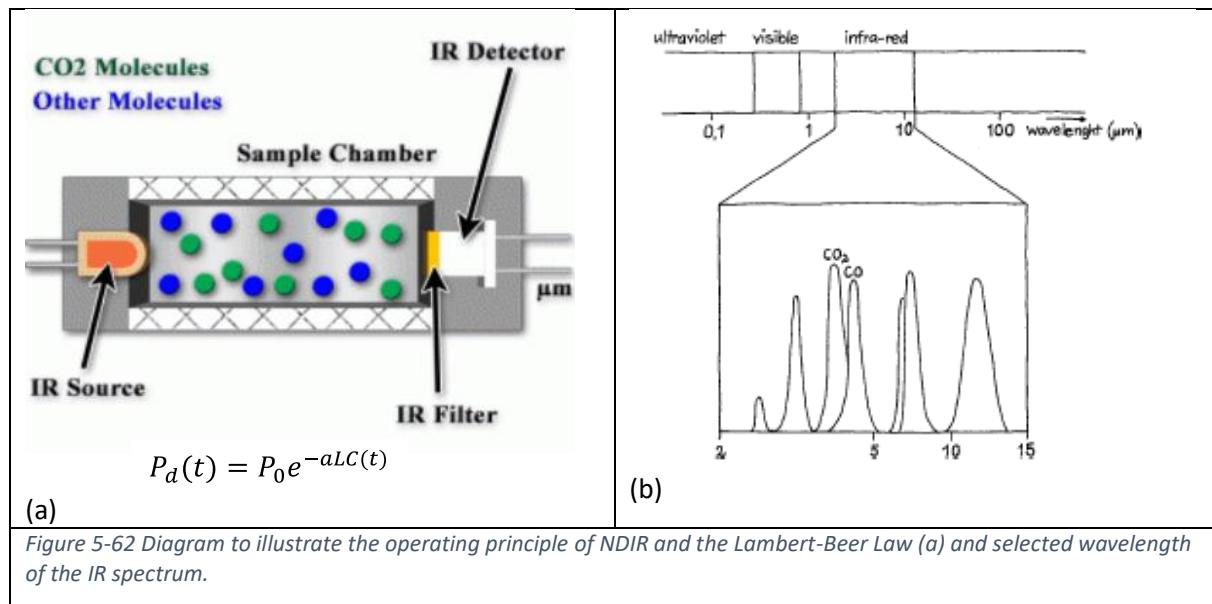


Figure 5-62 Diagram to illustrate the operating principle of NDIR and the Lambert-Beer Law (a) and selected wavelength of the IR spectrum.

Capnographs measure CO_2 partial pressure using the absorption of infra-red radiation at specific wavelengths. The infra-red wavelength should be specific for the gas that is measured. There are many gases used for anesthesia which limit the choice in wavelength for the sensor. For carbon dioxide a wavelength of 4.2 μm was selected (see Figure 5-62 (b)), it is in a vibration band of the IR spectrum of CO_2 . From a technical point this is a very difficult region of the IR spectrum. There are no LED light sources or suitable photo-diode detectors. For the IR source shown in Figure 5-62 a special tungsten filament lamp or gas discharge lamp with CO_2 in the gas is used. The IR light is transmitted into a gas flow cell and is absorbed by CO_2 gas in the cell. An IR filter absorbs the undesired

frequencies of the continuous IR spectrum. The IR light is detected by a thermopile or other sensitive thermo-electric detector. The absorption follows the Lambert-Beer law indicated in Figure 5-62 (a) and the concentration can in principle be determined from this relation [Webster]. However, the intensity of emitted light, geometry factors and the detector efficiency are not well known. These factors can be calibrated out by using a reference cell with a known concentration of CO₂. A concept for a differential NDIR cell is shown in Figure 5-63. Note that the same IR source and detector are used for detection of light transmitted through the sample and reference cells. In some devices the IR light is modulated to enable synchronous detection techniques. In the embodiment (b) a beam splitter and two-detector and two-optical filter solution is used to calibrate out intensity variations in the light source. One of the optical filters transmits CO₂ IR, the other a reference IR wavelength.

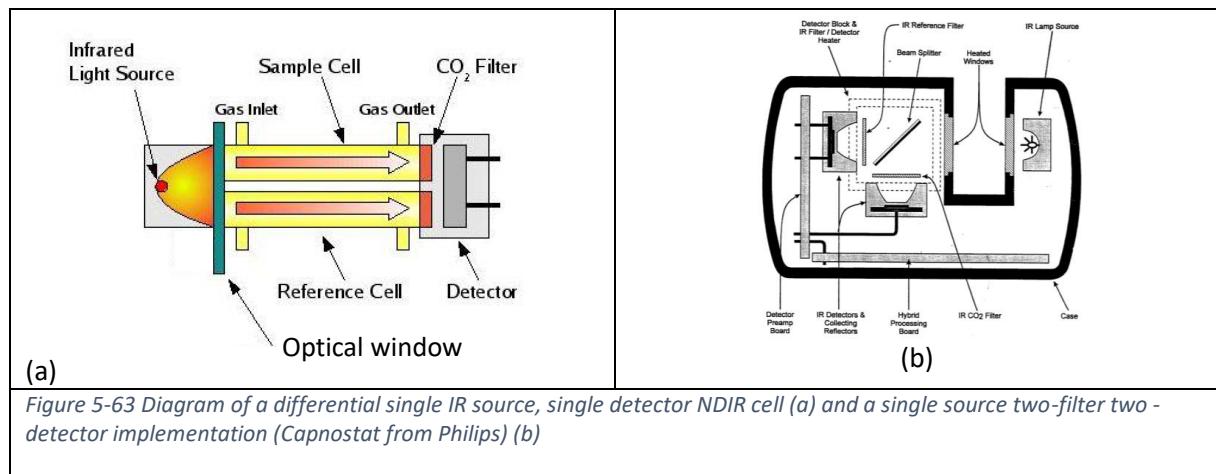


Figure 5-63 Diagram of a differential single IR source, single detector NDIR cell (a) and a single source two-filter two-detector implementation (Capnostat from Philips) (b)

There are mainstream and side stream Capnographs. A diagram of a mainstream capnograph from Philips (Capnostat) is shown in Figure 5-64 (a). In a mainstream embodiment inspired and expired air is measured directly. Air flows through a tube which is connected to an endotracheal tube or face mask. This system is used mostly during surgery and in the intensive care. Masks are not used because of air leaks and monitoring results are less reliable. The system is small and gives the most reliable capnography signals. In the Philips Capnostat the sample cell has a dual function, besides the capnograph function it is also a spirometer, and this enables volume capnography (i.e. plot of CO₂ partial pressure versus expired volume)

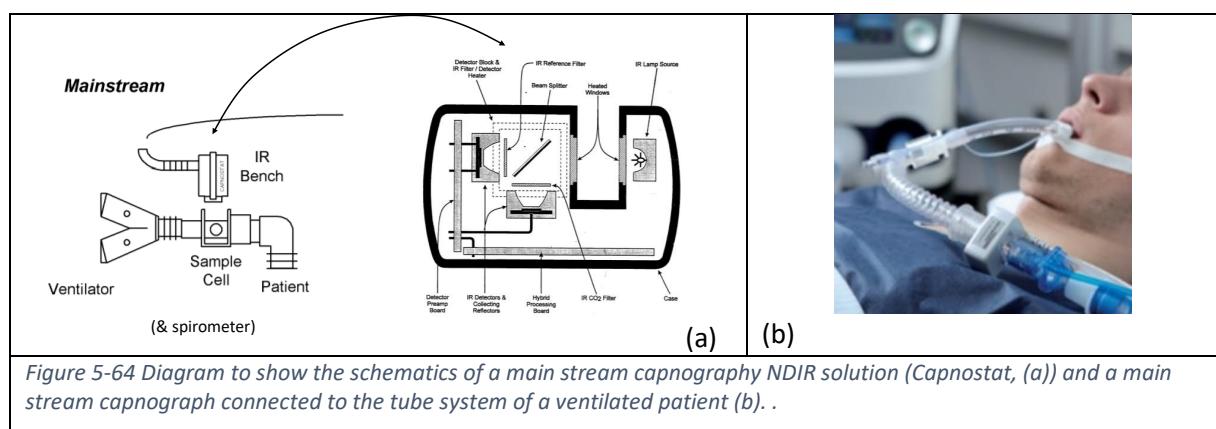


Figure 5-64 Diagram to show the schematics of a main stream capnography NDIR solution (Capnostat, (a)) and a main stream capnograph connected to the tube system of a ventilated patient (b).

A second embodiment is the side-stream capnography. This technique is illustrated in Figure 5-65 (a). Air is pumped from the patient attachment to a sensor system via water trap (expired air contains water vapor that condenses in the tubing). The pump and sensor are in a module that is integrated in the monitor. About 50 ml per minute is pumped to the sensor unit. The advantage is that capnography can be measured on both non-intubated and intubated patients. A disadvantage is the delay of a few seconds between the actual and measured values. Furthermore, condensation of water vapor in the tubing is a worry. Special measures are taken to limit or avoid water vapor condensation.

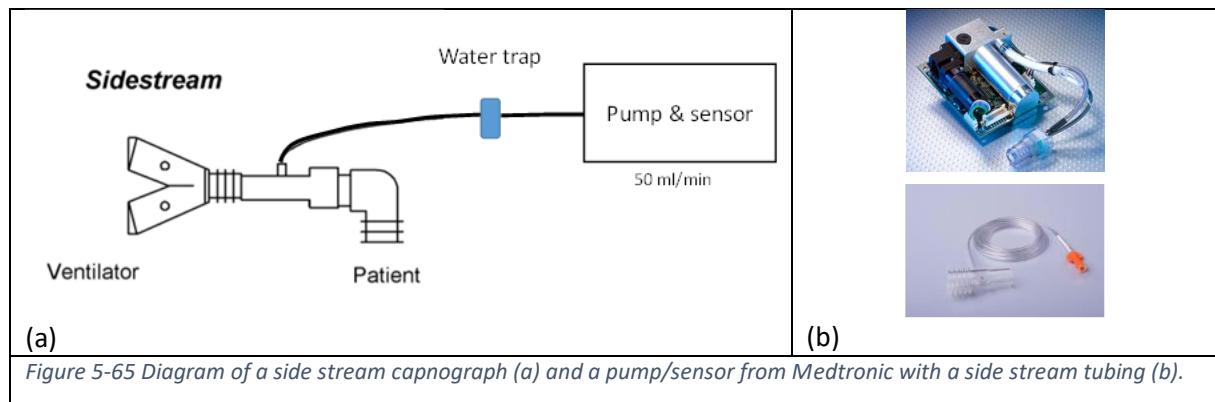


Figure 5-65 Diagram of a side stream capnograph (a) and a pump/sensor from Medtronic with a side stream tubing (b).

5.11 Pulse Oximetry

The oximeter device was developed to measure the relative concentration of oxygenated hemoglobin in blood. The absorption of light in a test tube filled with blood is used to explain the basic principle. The absorption of light in such a test tube can be described by the Lambert-Beer Law. The incident light intensity I_0 is partly absorbed by the blood and intensity I_d reaches the detector. The absorption scales exponentially with the length L , the concentration C of the specific molecule and the molar extinction coefficient ε . Note that ε is specific for a molecule and depends on the wavelength λ of the light. The molar extinction coefficients for oxygenated hemoglobin (HbO_2) and deoxygenated hemoglobin (Hb) are shown in Figure 5-68.

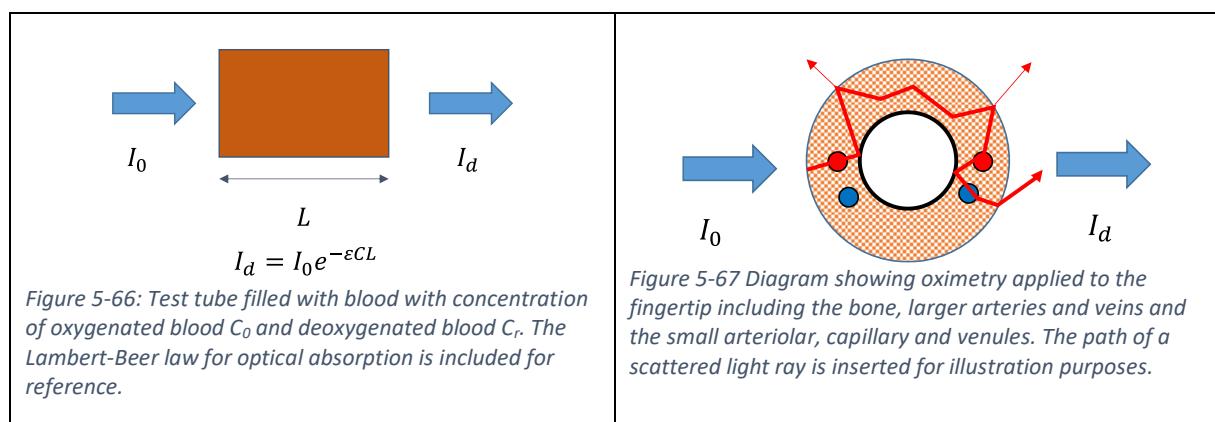


Figure 5-66: Test tube filled with blood with concentration of oxygenated blood C_o and deoxygenated blood C_r . The Lambert-Beer law for optical absorption is included for reference.

Figure 5-67 Diagram showing oximetry applied to the fingertip including the bone, larger arteries and veins and the small arteriolar, capillary and venules. The path of a scattered light ray is inserted for illustration purposes.

For small absorption (or extinction) the Lambert-Beer Law can be written as:

	$I_d/I_0 = e^{-\varepsilon CL} \approx 1 - \varepsilon CL$	5.11-1
	$E = 1 - I_d/I_0 = \varepsilon CL$	5.11-2

The extinction E is defined in equation 5.11-2. We assume there is absorption of light by HbO₂ and Hb only and that the molar extinction coefficients ε are wavelength dependent we can write equation 5.11-2 as:

	$E(\lambda) = \varepsilon_O(\lambda)LC_O + \varepsilon_r(\lambda)LC_r$	5.11-3
--	--	--------

C_O is the concentration of oxygenated Hb and C_r is the concentration of reduced or deoxygenated hemoglobin. Introducing the relative concentrations C_{OR} , C_{rR} and the total Hb concentration C_T :

	$C_{OR} = \frac{C_O}{C_O + C_r} = \frac{C_O}{C_T}, C_{rR} = \frac{C_r}{C_T}, C_T = C_O + C_r$	5.11-4
--	---	--------

it is now possible to write equation 5.11-3 as:

	$E(\lambda) = LC_T (\varepsilon_O(\lambda)C_{OR} + \varepsilon_r(\lambda)C_{rR})$	5.11-5
--	---	--------

The saturation of oxygenated hemoglobin is defined as:

	$SO_2 = \frac{C_O}{C_O + C_r} * 100\% = C_{OR} * 100\%$	5.11-6
--	---	--------

The product of the factor LC_T and the relative concentrations of oxygenated and reduced hemoglobin can be obtained by measuring the extinction E (equation 5.11-5) at two different wavelengths, typically a red and infrared wavelength are chosen. The saturation SO_2 can be obtained from a modified form of equation 5.11-6 by inserting the known terms $LC_T C_{O,r}$ instead of the concentrations $C_{O,r}$. Alternatively, the relation $C_{OR} + C_{rR} = 1$ can be used and equation 5.11-5 can be written as:

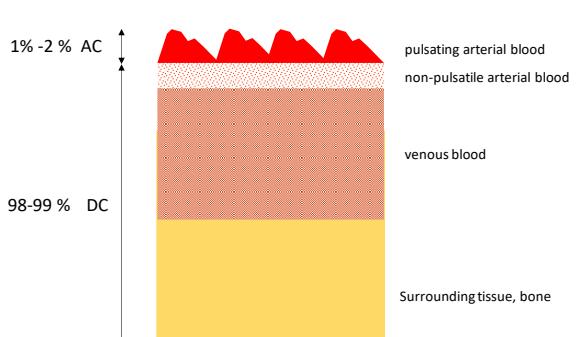
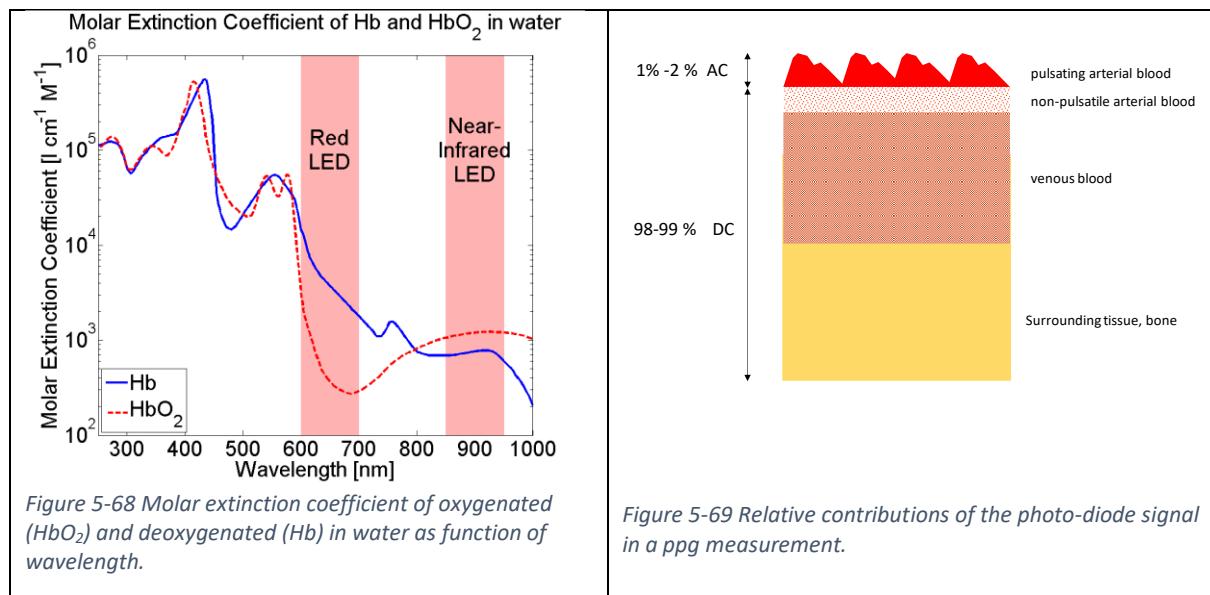
	$E(\lambda) = LC_T ((\varepsilon_O(\lambda) - \varepsilon_r(\lambda)).C_{OR} + \varepsilon_r(\lambda))$	5.11-7
--	---	--------

There are two unknowns (C_{OR} , LC_T) and by measuring at two wavelengths the Hb saturation with oxygen SO_2 can be determined from equation 5.11-7. Hence the saturation can be determined without the knowledge of the total Hb concentration and path length. By measuring at more than two wavelengths the total Hb concentration and concentrations of other molecules that absorb light can be determined. Oximeters with 8 or even more wavelengths are available, they are typically used for blood gas analysis in the laboratory environment. The blood can contain other forms of hemoglobin. Carboxyhemoglobin (carbon monoxide bound to hemoglobin, related to smoking, air pollution 10% - 45 %) and methemoglobin (an oxidized form of hemoglobin that cannot carry O₂) can be present in larger amounts (in case of a rare disease in which methemoglobin is not removed from blood values up to 70% are possible), cannot bind oxygen and can disturb the two wavelength measurement of SO_2 . Especially methemoglobin has a large absorption in the red and infra-red bands. The concentrations of these molecules can be detected when measurements are done at more wavelengths.

When placed on a limb (arm, leg) oximeters can measure the saturation of Hb in blood, but a value corresponding of a mixture of venous and arterial blood is found. When blood samples from an artery or vein are used in the laboratory this is not an issue. One solution to measure the real time oxygenation of Hb in arterial or venous blood is to place an oximeter at the tip of a catheter and measure the saturation of Hb in a large artery or vein. This is an invasive and obtrusive method. There is a need for a non-invasive measurement of the saturation of arterial blood. There is a good solution that is discussed in detail in the following section.

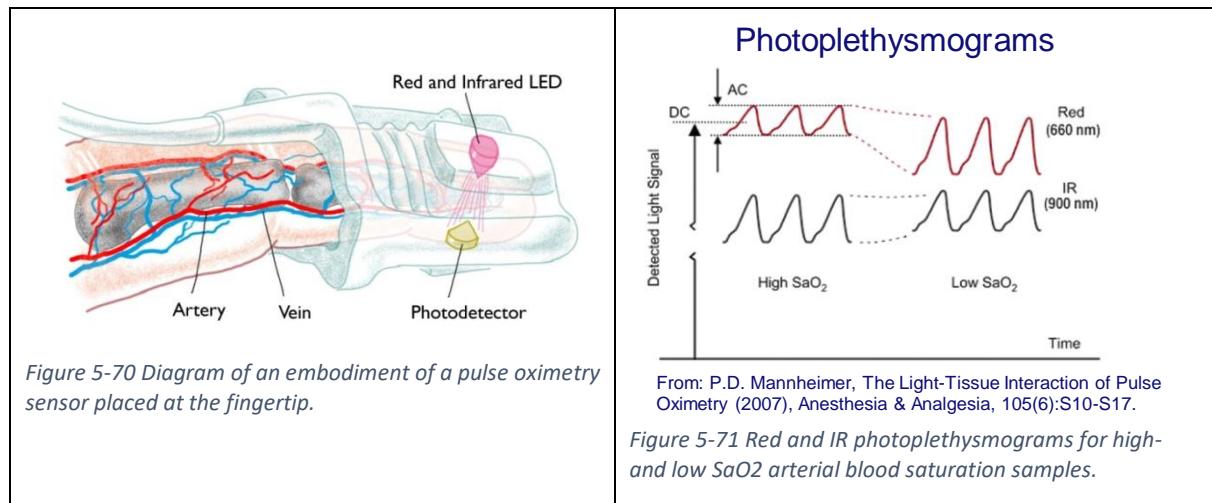
A non-invasive measurement of the oxygenation of arterial blood of a patient must be available in a multi-parameter patient monitor. It exploits the pulsatile nature of blood volume in the artery. The so-called pulse-oximetry method was invented in the 1970's by the Japanese bioengineer Takuo Aoyagi during a study of methods to measure cardiac output. The technique is based on the

following assumptions. The first is that oxygenated hemoglobin (HbO_2) and deoxygenated hemoglobin (Hb) have different molar extinction coefficients in the red and infra-red wavelength regions (see Figure 5-68). The extinction coefficient is smaller for HbO_2 at the red wavelength and larger than that of Hb at the IR wavelength. By measuring the pulsatile ppg signal at two (or more) wavelengths, one in the red region and one in the IR region it should be possible to estimate the ratio of the concentrations of the two forms of Hb in arterial blood.



The main assumption is that the pulsatile ppg signal is due to arterial blood only. By measuring the small pulsatile part of the ppg signals of the red and IR signals the fraction of oxygenated Hb ($\text{HbO}_2/(\text{Hb}+\text{HbO}_2)$) of arterial blood can be estimated. This fraction is called SpO₂ or percentage saturation of hemoglobin. It is important to realize that pulse oximetry does not measure the oxygen concentration, for this parameter the total Hb concentration must be known as well as SpO₂. Other assumptions are that Hb and HbO₂ are the only oxygen carrying species and that the optical path length of the arterial blood is the same for the two optical paths. The percentage oxygenated blood can also be measured in an arterial blood sample in the laboratory. This parameter is called SaO₂, i.e. percentage saturation of arterial blood.

In Figure 5-70 a diagram of a fingertip pulse oximeter sensor is shown. The fingertip is one of the best locations to measure SpO₂, it has a dense network of small arteries and has a large optical transmission and arterial blood volume variation. The sensor consists of two LEDs, one red (~650 nm) and one IR (~940 nm) LED that are modulated such that red and IR red can be detected separately at the silicon photo detector and out of band noise is suppressed. The light sources and detector are located at opposite sides and are enclosed in a rubber elastic casing that presses the two LEDs and photodetector with a controlled contact force to the skin. Besides fixating the LED and photodetector the casing also shields the sensor for ambient light which a very important feature of this embodiment. Ambient light can influence the measurement quality to a large degree.



In Figure 5-71 measured red and IR ppg signals are shown for high and low SaO_2 cases. The ppg signal consists of a pulsatile ppg component that is marked as AC signal and an average signal that is marked as DC signal. Both the DC and AC signals of red and IR vary with the Hb saturation of arterial blood. SpO_2 is determined from the ratio R which is defined as:

$$R = \frac{AC(\text{rd})/DC(\text{rd})}{AC(\text{IR})/DC(\text{IR})} \quad 5.11-8$$

The relation between the ratio R and SpO_2 can be derived from the Lambert-Beer law. A model prediction of this relation is shown in Figure 5-72. There is a large deviation between the model and the measured values. It was found that besides absorption, scattering of light on tissues and bone and different optical paths for red and IR light needed to be taken into account (see Figure 5-67). In practice a look-up table or a polynomial fit model is used to relate SpO_2 to the measured R value. The calibration is done on a volunteer group using measured blood gas data. Note that this calibration table is only valid for the specific sensor type at the tested body location. Placing the sensor on another body location can lead to large errors. Typically, SpO_2 can be measured with a few percent accuracy (2%-3%) between 80% and 100% saturation. For lower saturation the error margin becomes very large. The accuracy of the SpO_2 parameter is improved by post-processing algorithms that remove artifacts from the measured signals.

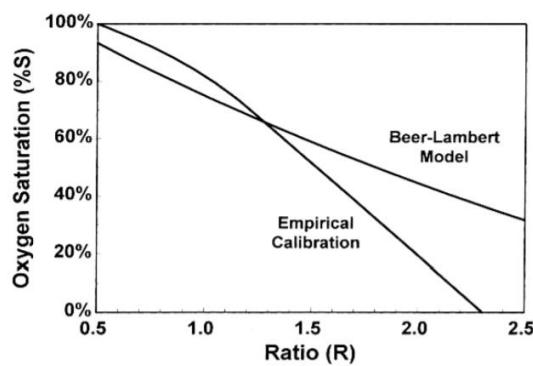


Figure 5-72 Measured and model based relation (Lambert-Beer model) between SpO_2 and R

There are only a few body locations which are suited for high quality pulse oximetry measurements. These are shown in Figure 5-73. There is the group of peripheral sensors which include the fingertips

and toes and the earlobe. These locations are convenient for user and patient and offer the best signals and lowest signal to noise. However, during blood flow centralization as occurs for instance during shock states, trauma and myocardial infarction the sensors are useless as blood flow to the peripheral areas is blocked. There are two locations that remain perfused during blood flow centralization. These are locations that receive blood from the arterial system in the brain that remains perfused under all conditions. These are the forehead and alar wing locations. The alar wing site sensor performs excellent under all type of clinical conditions. Note again that these sensors must be used only for the specific location. These sensors are more obtrusive for the patient.

Peripheral Sensors

Ear lobe



fingertip
toes



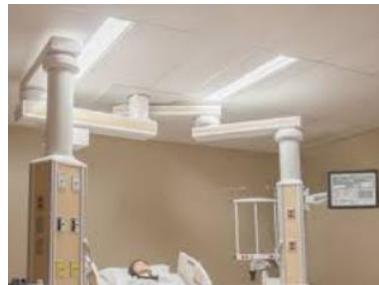
Figure 5-73 Body locations for high quality pulse oximetry measurements.

At first sight the pulse oximetry measurement seems to be simple. This is misleading. There is large variation in transmission values depending on skin color, nail polish, finger size (more than a factor 100 variation in the transmission level). There is also a large variation in the relative AC amplitude, it can vary by more than two orders of magnitude. The dynamic range of the measurement signals is very large. The measurement is also sensitive to noise. Both LED driver noise and electronics noise needs to be reduced to 100dB. Furthermore, the impact of environmental factors such as stray light and motion on measurement accuracy is large (see Figure 5-74).

Sunlight



Ambient lighting



LED lighting



Figure 5-74 Environmental lighting conditions that degrade pulse oximetry accuracy.

The red and IR LED light is modulated, both QAM and time domain modulation is used. Duty cycles and frequencies vary per manufacturer, but all are selected such that interference with ambient lighting is minimal. Unfortunately, with the introduction of LED lighting this has become a major challenge as each LED lighting supplier uses other modulation schemes. This will become a challenge for future pulse oximetry systems in hospitals.

As mentioned earlier the post processing algorithms to clean the signals of noise and artifacts are of extreme importance for the measurement quality. Two important post processing algorithms are the Philips FAST algorithm and Massimo's SET algorithm. Mild and repetitive motion artifacts can be removed, severe motion artifacts cannot be removed. The measurement detects this condition and warns the user. These algorithms are important selling features for customers.

5.12 References

[Webster] John G. Webster, Medical Instrumentation, Application and Design, John Wiley and Sons, fourth edition, ISBN 9780471676003.

[Blom] J.A. Blom, Monitoring of Respiration and Circulation, CRC Press LLC, ISBN 9780849320835

5.13 Questions

1. Describe the different parts of a transducer system
2. What are the equations that link the input and output signals of zero order, first order and second order systems? What are the important parameters.
3. How can a second order system be designed to avoid ringing effects?
4. What is measurement accuracy, precision and resolution?
5. What are the outputs of a clinical measurement?
6. Describe how a strain gage pressure transducers measures pressure.
7. Describe the pneumotachograph and how it can measure gas flow.
8. Derive the expressions for the Venturi and Pitot tubes
9. Why is the height of a rotameter's bobbin related to the flow in such a complex way?
10. Why can a specific rotameter only be used for a specific gas?
11. Describe the hot wire anemometer.
12. Discuss how blood velocities are measured using the continuous ultrasound technique.
When is this technique especially useful?
13. Discuss how the blood flow in a vessel can be measured using the pulsed ultrasound technique.
14. Which tissue depth limitations are important in the pulsed ultrasound technique? Why?
15. Explain how an acoustic pulse that is sent by a multi-element pulsed Doppler transmitter array can be given a varying spatial direction.
16. What is plethysmography?
17. Describe the cuff plethysmography sensor.
18. What is photo plethysmography? Describe the operating principles?
19. Describe the schematics and electronic parts in a ppg system.
20. Why is the ppg sensor sensitive for motion artifacts?
21. Describe why bio-impedance can be used for plethysmography measurements?
22. Describe the transducer principle and frequency dependence of the contact resistance of an Ag/AgCl electrode
23. What is an ECG signal and how can it be measured?
24. Describe the main components and schematics of an ECG measurement system
25. Describe the different electrical signals that are measured during ECG, what is the relative magnitude and how can the ECG signal be separated from the other signals?
26. Which body temperature is of most interest to a clinician and why?
27. Describe the main temperature sensing technologies
28. Describe non-dispersive infrared absorption spectroscopy.
29. What is a capnograph, how can it be made specific for CO₂?
30. What is the difference between main stream and side stream capnography? What are the pro and cons of these methods?
31. How can overlapping spectra of different gases be handled?
32. Describe how, in the oximeter, Beer's law and the known properties of the absorption spectra of hemoglobin and oxyhemoglobin can be used to determine the oxygen saturation of blood. Why are measurements at (at least) two wavelengths necessary?
33. How does a pulse oximeter differ from an oximeter?
34. What is the effect of a significant presence of methemoglobin on the readout of a pulse oximeter?
35. Which measurements are needed if the total concentration of oxygen in blood needs to be determined?
36. Explain the differences in DC and AC signals in Figure 5-71 for blood with low and high SaO₂.

37. Explain the trend in magnitude with increasing R in the empirical SpO₂-R curve in Figure 5-72.

6 Measurements of the Respiratory System

6.1 Introduction

The main function of the respiratory system is oxygenation of arterial blood and removal of excess carbon dioxide from venous blood. In short this involves air flowing into and out of the lung and gas exchange between the alveoli and capillaries. During a disease these functions may be impaired and patients are suffering from dyspnea. For the diagnosis and treatment clinicians want to gather data of the main functions of the respiratory system. In this chapter the main measurements of the respiratory system (module 4 of the course) are described. Measurements of the pulmonary system are needed for diagnosis, control of treatment and therapy and to monitor the effects of medicines.

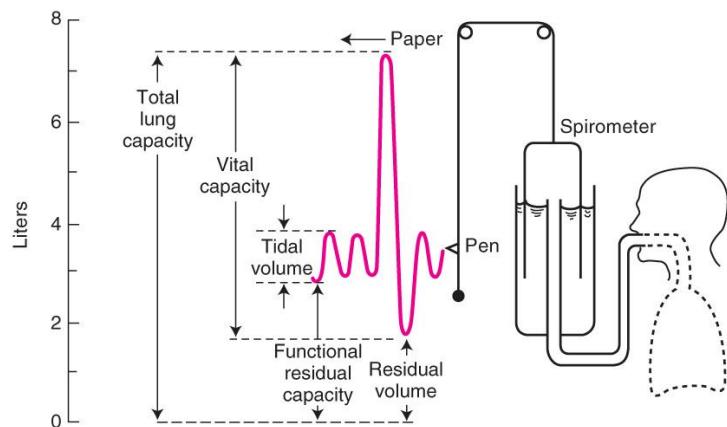
6.2 Measurements

In restrictive lung diseases the lung volume and/or elasticity of the lung tissue is affected. Patients have difficulty in expanding the lung volume. In obstructive lung diseases exhalation of air from the lungs is limited. Measurements have been developed to distinguish between these types of lung disease and monitor the status and severity of the disease.

In lung function measurements lung volumes and changes in lung volume are studied (see **Error! Reference source not found.**). These tests include effort dependent inspiration and expiration tests. Furthermore the measurement of dead space regions in the lung is of importance. Dead space is the volume of the respiratory system where there is no gas exchange with blood. Air flow into and out of the lung and airway resistance are important measurements for the obstructive disease types. Measurements of oxygen diffusion and measurement of CO₂ partial pressures in expired air offer insight in gas exchange between the lung air and blood in the lung. Finally oxygenation of blood and so-called blood gases important require sampling of blood, this is done typically during high levels of care. These measurements fall outside the scope of this course. A short summary of these tests is available in the appendix. Measurements are done in lung function laboratories, home monitoring is possible for the simplest measurements. Some of the measurements are used when patients are monitored during mechanical ventilation. The most important measurements for the lung function are discussed in the following sections. We will start with the measurement of lung volume(s).

6.3 Lung Volumes

One of the most common tests is the measurement of the static lung volumes. The most important lung volumes are shown in **Error! Reference source not found.**. During normal breathing the tidal volume (TV) is inhaled/exhaled. The lung volume after a maximal inhalation is the total lung capacity (TLC). The lung volume after a maximal expiration is the residual volume (RV). This volume cannot be expired. The lung volume after a normal breathing effort is the functional residual capacity (FRC). The maximum volume that can be exhaled is the vital capacity (VC). The maximum volume that can be expired from the FRC level is the expiratory reserve volume (ERV). Note that most lung volume parameters include the residual volume (RV), this is a volume that cannot be expired. These volumes cannot be measured with a spirometer. These volumes are very important for diagnosis and therapy, dedicated measurements have been developed. Changes in lung volume can be measured easily, **Error! Reference source not found.** shows a spirometer device which is most often used for measurements of volume changes. This device is used widely and is discussed in the next section.



6.3.1 Lung volume changes - Spirometer

In the lung function laboratory, the various lung volumes are usually measured by having the patient breathe into and from a spirometer. The principle of spirometry is explained in Figure 6-1. The spirometer-lung is a closed system. The extra, highly compliant volume V_s of the spirometer is added to the lung volume V_L . Since the total volume $V_s + V_L$ is constant, changes in V_L will reflect changes of V_s , which can be measured. The spirometer is based on this principle. Figure 6-2 shows a schematic diagram of a spirometer. A lightweight airtight clock, whose mass is compensated for with a counterweight, can freely move up and down. A water lock ensures that the movement encounters negligible resistance. The pressure inside the clock is equal to the atmospheric pressure. A carbon dioxide absorber ensures that the patient will not rebreathe his expired CO₂. The CO₂ is chemically bound in the form of calcium carbonate.

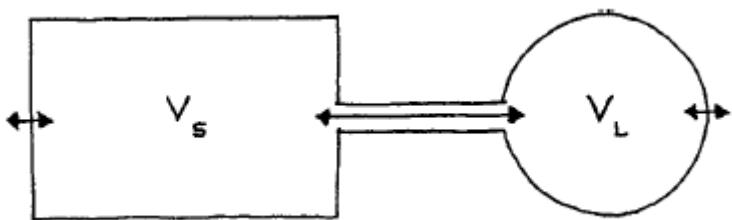


Figure 6-1 During spirometry, the patient breathes back and forth into the spirometer, whose volume V_s is measured. The pressure inside the spirometer is maintained at atmospheric pressure.

In order to minimize the added dead space, two tubes, each with a unidirectional valve, connect the patient to the spirometer. A thermometer measures the temperature of the inspired gas; the expired gas is assumed to be at 37 °C.

The various volumes can be measured by asking the patient to perform maneuvers, an example is a maximum in- or expiration effort. Figure 6-3 shows a spirometer recording. Initially, the patient breathes quietly from the FRC volume; volume changes reflect the tidal volume (TV). Since the patient consumes oxygen, the spirometer's volume will decrease slowly over time. This decrease, which is visible as a slow drift superimposed on the volume changes due to normal breathing, is the patient's oxygen consumption per unit of time. It is of the order of 250 ml/min. When the patient is asked to exhale maximally the ERV volume can be measured. When the patient is asked to inspire from FRC to the maximum lung volume the IRV volume can be measured. The vital capacity can be measured when the patient first expires to RV and then inhales to TOC. This volume is called vital

capacity VC. Figure 6-3 also shows some other maneuvers. The FEV and FEV_1 (Forced Expiratory Volume after 1 second) test and the FIV and FIV_1 (Forced Inspiratory Volume after 1 second) tests. These tests are very important and provide measures for the power that the respiratory muscles can develop and for existing restrictions or obstructions. These tests are discussed later in more detail. Finally the maximum total volume that can be inspired for a period of 1 minute can be measured.

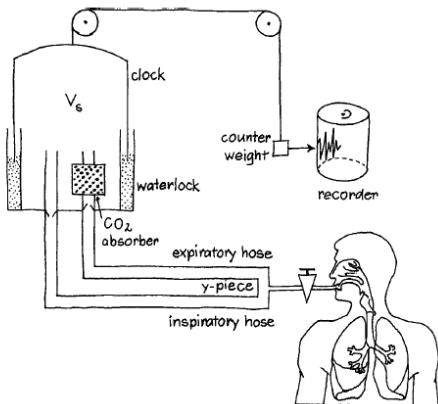


Figure 6-2 A spirometer is an airtight counterbalanced clock, whose volume is measured. Unidirectional valves ensure that minimal dead space is added. A CO_2 absorber removes CO_2 from the expiratory gas.

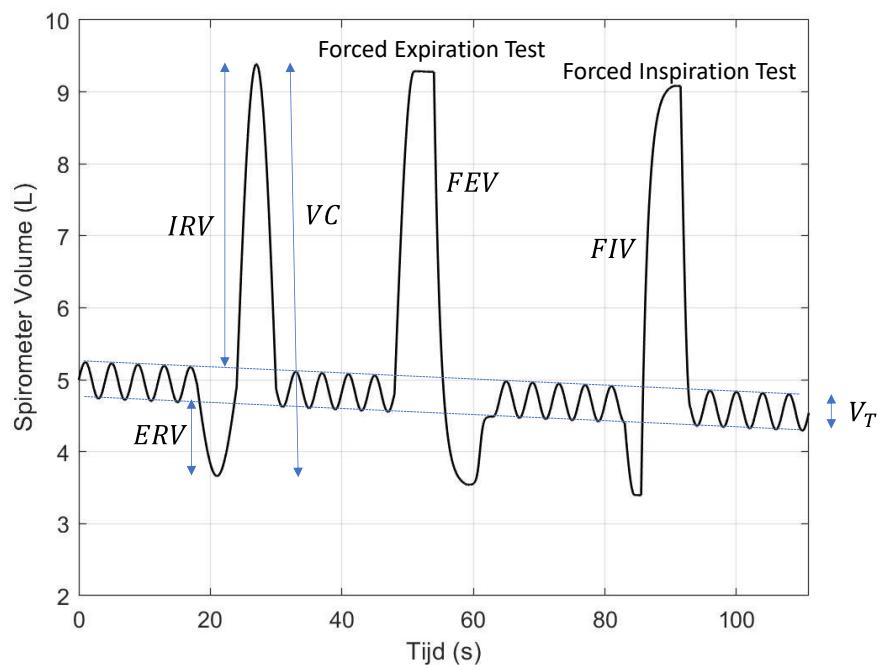


Figure 6-3 When the patient performs respiratory maneuvers, the spirometer's volume changes reflect lung volume changes. The slow drift in the recording corresponds with oxygen consumption.

6.3.1.1 Corrections for temperature, pressure and water vapor partial pressure

To accurately interpret and compare spirometry results, several corrections or compensations are necessary. When the gas moves back and forth between lungs and spirometer, its pressure, composition, temperature and water vapor content vary. In slow maneuvers, the pressure inside the lung (the alveolar pressure) can be assumed to equal atmospheric pressure; the flow is small, and the pressure drop over the airway resistance is small too. In rapid maneuvers, where high flow rates

occur, this is not true anymore; in these cases, an esophageal balloon measurement can establish the alveolar pressure, if required. The temperature of the gas in the spirometer is measured; the temperature of the gas in the lungs can be assumed to be 37 °C. The expired gas is water vapor saturated. Some of this water vapor condenses on its way to the spirometer's lower temperature, but inside the spirometer the gas will still be water vapor saturated, but *at a different temperature*.

First, compensation for water vapor content is required. By subtracting the strongly temperature dependent water vapor partial pressure P_{H_2O} from the gas pressure P_{tot} , we obtain the "dry gas" partial pressure:

$$P_{dry-gas} = P_{tot} - P_{H_2O} \quad 6.3-1$$

We can use Boyle's law ($P V / T = \text{constant}$) for dry air and obtain compensations for temperature and pressure. The result is equation 6.3-2, which converts a gas volume V' from a temperature T' and pressure P' to an equivalent volume V at temperature T and pressure P :

$$V(T) = V(T') \frac{P' - P_{H_2O}(T')}{P_{tot} - P_{H_2O}(T)} \cdot \frac{T}{T'} \quad 6.3-2$$



(a) Historical Spirometer



(b) Modern Spirometer



(c) USB coupled device



(d) Benchtop device

Figure 6-4 Historical (a) and modern embodiments of spirometer systems in a lung function laboratory (b), USB coupled device for home monitoring (c), (d) benchtop device with direct read out during a spirometer test.

For modern measurements of the lung volume the drum spirometer is replaced by smaller and more intelligent devices which measure flow (Figure 6-4 b,c,d) and calculate volumes and volume changes by integration of flow rates and can show data and test reports directly on a display results (d). For measurements at a general practitioner benchtop devices are available (d). For use in the home, USB coupled PC based systems or wireless connected devices are available (c). In figure Figure 6-4 (d) a spirometer test on a benchtop device is shown.

In the following descriptions of spirometry tests, we will assume that these pressure and temperature compensations will be performed. In the figures, we will also use the basic diagram of Figure 6-1 rather than the more complex Figure 6-2 in order to present the measurement principles more clearly.

6.4 Measuring absolute lung volumes

The residual volume RV cannot be voluntarily exhaled; the residual volume, functional residual capacity FRC and the total lung capacity TLC cannot be measured directly. The determination of these *absolute* lung volumes requires a different method. Three methods are in clinical use, the nitrogen washout method, the helium dilution method, and body plethysmography.

6.4.1 Nitrogen Wash-Out method

In the nitrogen washout method, the gas in the lungs-and thus the nitrogen in the lungs is "washed out" into the spirometer (Figure 6-5). Initially, the spirometer is empty or contains a gas (mixture) in which no N_2 is present. During about 10 minutes or until the expired N_2 concentration is lower than 2%, the patient inhales pure oxygen and exhales into the spirometer. After this period, practically no N_2 is present in the lungs anymore. The reason is the alveolar membrane's low N_2 diffusion capacity; although a great deal of N_2 is in solution in the tissues, very little of it crosses into the lungs during a time as short as 10 minutes.

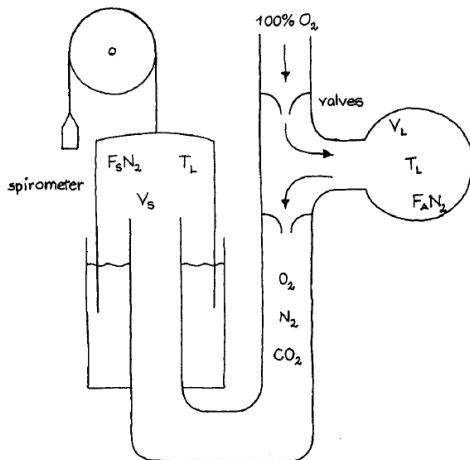


Figure 6-5 Inspiratory gas is 100% oxygen. During the test, the lung gases are expired into the spirometer. The final N_2 - mass in the spirometer is equal to the initial N_2 - mass in the lung.

At the end of the test period, the N_2 concentration in the spirometer is measured. When the test starts, the normal concentration of N_2 in the alveoli is determined from the end-expiratory N_2 concentration; at the end of expiration, gas that comes from the alveoli is exhaled. Since the N_2 mass in the volume under consideration has not changed, we can compute the lung's total volume:

$$C_{N_2}(t_{start}) \cdot V_{lung} = C_{N_2}(t_{end}) \cdot V_s \quad 6.4-1$$

If the FRC is to be measured (which is normally the case), the test starts and ends at the end of a normal expiration. The measurement of N_2 concentration during the test gives information on fast and slow regions in the lung. This is shown in Figure 6-6. For a normal lung the concentration scales down exponentially with the number of breaths, i.e. a straight line in a semi-logarithmic plot. In a normal lung with a larger area of slow alveolar ventilation a more complex curve-linear behavior is seen in such a plot.

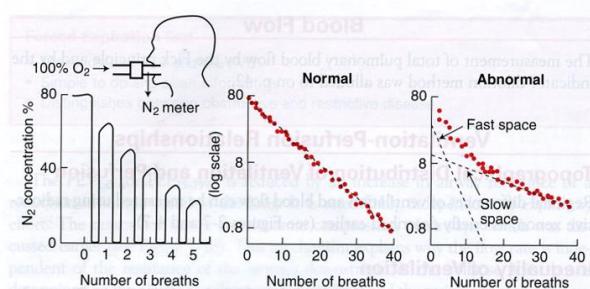


Figure 6-6 N_2 concentration as function of the number of breaths. Abnormal behavior shows up as deviations from a straight line in a log plot.

6.4.2 The Helium dilution method

In the helium dilution method (Figure 6-7), the patient breathes back and forth into the spirometer. Initially, the spirometer contains a known (measured) concentration of helium. After a period of about 10 minutes or until the spirometer's He concentration does not change anymore, the helium has been redistributed over the larger volume of spirometer plus lungs combined and its concentration is measured again. Due to the alveolar membrane's low helium diffusion capacity, practically no Helium has crossed from lung to blood and tissues during the test interval. Therefore, since the helium mass has not changed, its concentration in the spirometer can be measured and spirometer volume is known, the relationship 6.4-2 provides us with the total lung volume (V_{lung}).

$$C_{He1} \cdot V_{spir1} = C_{He2} \cdot (V_{spir2} + V_{lung})$$

6.4-2

Again, if the FRC is to be measured, the test starts and ends at the end of a normal expiration.

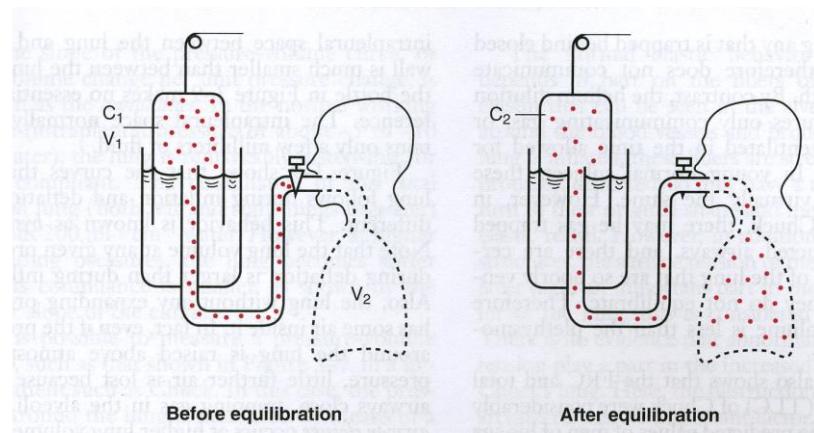


Figure 6-7 Initially helium is found only in the spirometer and in the hoses. At the end of the test, it has been redistributed over a larger volume, which now includes the lung volume.

6.4.3 Whole Body Plethysmography



Figure 6-8 Schematic diagram of a whole body plethysmograph system and actual device during test.

A body plethysmograph is essentially a spirometer large enough for the patient to sit in (Figure 6-8). Instead of the constant pressure varying volume of the spirometer, we now have a constant volume box whose pressure can be measured. The patient sits in the box, the door is closed and a waiting time for equilibration of air the temperature and air pressure in the box is needed. The patient

breathes air from the external environment via the spirometer. The most basic embodiment is discussed, practical devices are much more complex due to the calibration and extremely precise measurements that are needed. To determine the absolute lung volumes, the patient inhales or exhales to a particular volume (usually FRC), at which time a shutter closes the tube through which breathing takes place. When the patient exerts in- or expiratory effort against the closed shutter, the chest volume changes (the shift volume), the box volume changes with the same magnitude but opposite sign. The resulting pressure change in the box can be measured (see Figure 6-8). Note that there is no air flow, the amount of thoracic gas does not change and the lung pressure decreases (gas compressibility!). Since there is no flow, the lung and mouth pressures are equal.

By applying Boyle's law ($P V = \text{constant}$) to the box volume, the chest (or lung) volume change ΔV_{lung} is computed (6.4-3). This small change is called shift volume. It is of the order of 10 to 50 milliliter. This is orders of magnitude smaller than the box volume (~ 1000 liter) so very small pressure changes have to be measured. This is a difficult measurement as drifts and other noise sources need to be controlled and reduced.

$$V_{box} \cdot P_{box}(t_1) = (V_{box} + \Delta V_{lung}) \cdot P_{box}(t_2)$$

6.4-3

The box volume V_{box} is a parameter that can be determined by adding a calibration device that applies known changes in box volume and subsequent box pressure variations enable a direct calculation of the box volume minus the volume of the subject.

A second application of Boyle's law on the pressure and volume of the fixed amount of gas in the lung, before and during the closed shutter respiratory effort gives:

$$V_{lung}(t_1) \cdot P_{AWO}(t_1) = V_{lung}(t_2) \cdot P_{AWO}(t_2) = (V_{lung}(t_1) + \Delta V_{lung}) \cdot P_{AWO}(t_2)$$

6.4-4

In 6.4-4 $V_{lung}(t_1)$ is the only unknown, its value can be computed. Note that the volume $V_{lung}(t_1)$ also includes trapped gas volumes, i.e. regions in the lung that do not receive fresh air. If the patient's lung contains trapped gas volumes that cannot communicate with the respiratory tree, body plethysmography and spirometry based N₂ washout and He-dilution will give different results. Besides absolute lung volumes other lung function tests (spirometry) on volumes and effort are possible. Finally The measurement of airway resistance with the whole body plethysmograph is the gold standard measurement of airway resistance (section XXX). The body plethysmography method is the most accurate method to measure the residual lung volume (RV) including trapped gas regions of the lung, it also enables most other lung function tests.

6.5 Measuring dead space and unequal ventilation

Figure 6-9 shows the CO₂ concentration as a function of the expired gas volume. When an expiration starts, the first gas that is exhaled (region 1) comes from the anatomical dead space. This gas, inhaled in the previous breath, normally contains no CO₂. When the exhalation is well underway (region 3), the exhaled gas comes from the alveoli. In this region, the CO₂ concentration shows a slight increase. This *plateau slope* is due to the gravity induced inhomogeneous emptying of the alveoli. In region 2, we observe a mixture of dead space gas and alveolar gas; as time progresses, the conducting zone contributes less, and the respiratory zone contributes more to the expired gas. A numerical value for the dead space volume can be read from the curve as the volume where the CO₂ concentration reaches 50% of its final value, the end-expiratory CO₂ concentration.

The CO₂ concentration as a function of time is known as the capnogram. The expiratory part looks like Figure 6-9. During inspiration, the CO₂ concentration is normally zero. A period of a typical capnogram is shown in Figure 6-10.

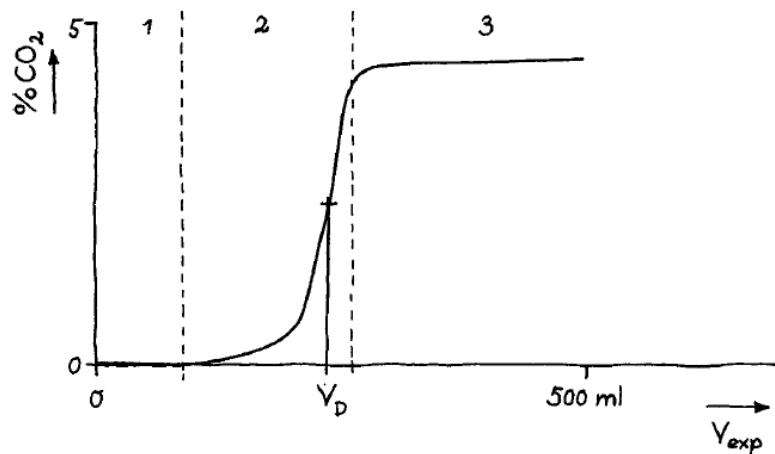


Figure 6-9 Expired concentration as a function of expired volume.

Since the dead space volume is exhaled first, pure alveolar gas will be exhaled only near the end of expiration. From this follows the general rule: if we want to know the composition of the alveolar gas, the *end-expiratory measurement* (ETCO₂: end-tidal CO₂ pressure) will reflect it. However, this will only be true if the expired volume is significantly larger than the dead space volume. The ETCO₂ parameter is a good measure of the arterial partial CO₂ pressure. It is used frequently during mechanical ventilation ventilation and during surgery to fine tune the ventilation minute volume. It is also an indirect measurement of metabolic activity and cardiac output.

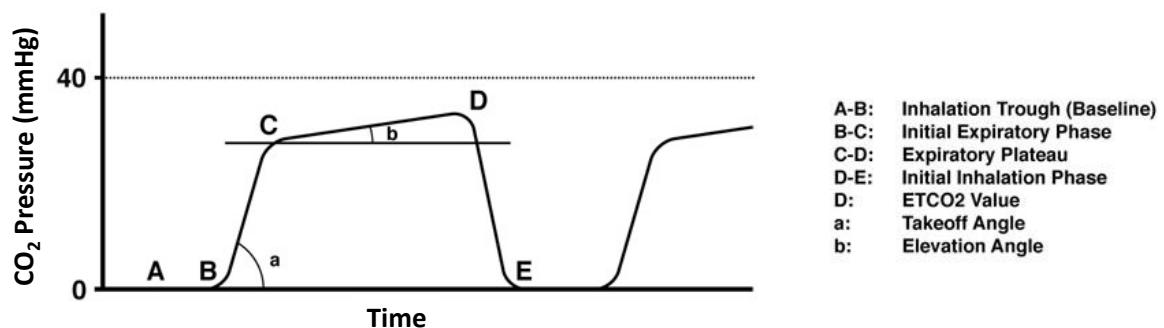


Figure 6-10 The expired CO₂ pressure as a function of time is called the capnogram. Important parameters are indicated in the figure. The dashed line is the arterial CO₂ pressure.

6.5.1 Single breath Nitrogen washout method

The single breath nitrogen washout curve, shown in Figure 6-11, is similar to the CO₂ washout curve of Figure 6-9. It is obtained by measuring the N₂ concentration during a maximal expiration after a single maximal inspiration of 100% O₂. The inspiration starts at the residual volume. Initially, N₂- free dead air gas (pure O₂) is exhaled. Next comes mixed gas (partially dead space O₂, partially alveolar gas containing N₂). Then there is, as was the case with the CO₂ washout curve, a sloping plateau; its explanation is again that the larger alveoli at the top of the lung expand less; they take up less O₂, and thus the gas that they return is less "diluted" by O₂. It contains a higher concentration of N₂. The curve of Figure 6-11 shows this.

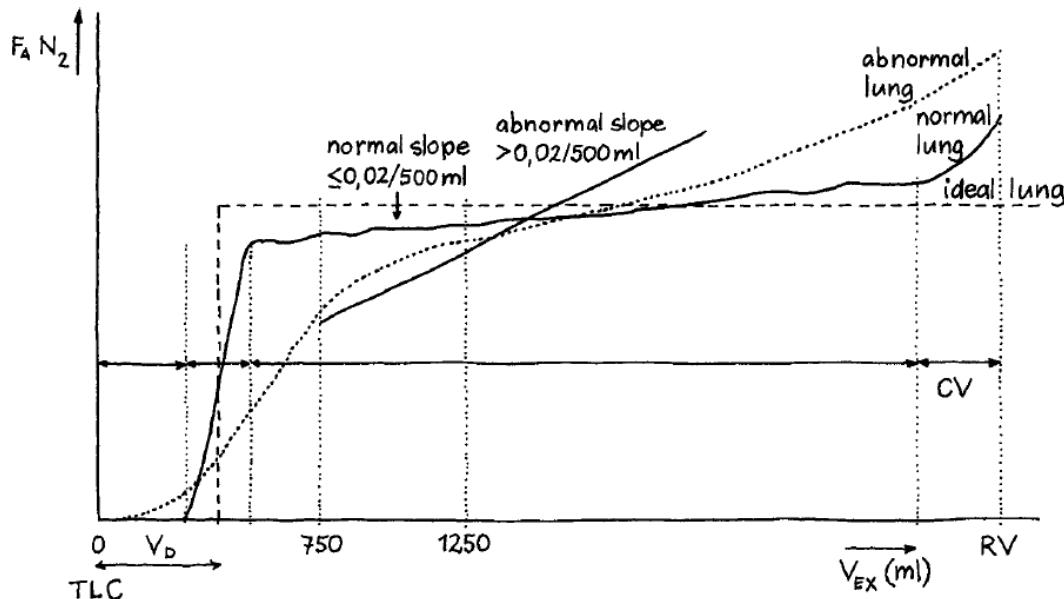


Figure 6-11 Expired N_2 concentration as a function of expired volume.

The plateau concentration can be calculated if we know the residual volume RV and the total lung volume TLC. If we assume that $RV = 1.2 \text{ l}$, it contains $0.8 * 1.2 \text{ l} N_2$. After inspiration of 100% O_2 , this N_2 is distributed over the whole lung volume. If $TLC = 6 \text{ l}$, $0.96 \text{ l} N_2$ means an N_2 fraction of $0.96/6$, i.e. a concentration of 16%.

The final part of the curve can only be seen in forceful exhalations, when the lower parts of the lung are fully collapsed. Only gas from the top of lung, with a large N_2 concentration, is exhaled here. This finally exhaled volume is called the *closing volume*. In extreme cases, when gas is exhaled from those parts of the lung that were not previously diluted by the 100% O_2 inspiration, the closing volume gas will contain about 80% N_2 .

Figure 6-11 shows that the curve is different in abnormal lungs, which have a significant physiological dead space. Not only is the dead space volume larger, the mixing zone is wider as well. Unequal ventilation is also visible in a greater slope of the plateau. An index for unequal ventilation is obtained by fitting a straight line through the N_2 concentrations that occur after 750 ml and 1250 ml have been exhaled. Slopes larger than 0.02 (2%) per 500 ml are considered abnormal.

6.6 Lung Compliance

The elastic properties of the lung are important for diagnosis of different lung diseases and are of importance during mechanical ventilation. The lung compliance is a measure for lung elasticity. The chest expansion during normal breathing is equal to the combined lung-chest wall compliance. For measurement of the lung compliance both the change in lung volume and the change in the transpulmonary pressure (i.e. the difference between alveolar air pressure and intra pleural space pressure) is required. A “simple” way to measure the lung compliance is having the patient breathe from a spirometer in volume increments of 500 ml while measuring the esophageal pressure. After each volume increase the lung needs a short no-flow stabilization period. It is essential that the patient's glottis remains open; only then will the alveolar pressure be zero. The measurement points, when connected by a smooth line, form a volume-pressure curve, as we already saw in other lectures. The *static lung compliance* is the slope of the curve. Figure 6-12 shows a set of volume-pressure curves; the lung compliance (the slope of a curve) can vary a great deal, depending upon

the physical properties of the lung tissues (normal, stiff or compliant lung). We also see that the curves are not straight lines, i.e. the volume-pressure relation is not linear. The lung compliance is not constant, although in the normal operating range and small tidal volumes it is nearly constant. In particular, the compliance decreases at high lung volumes. The figure also shows that the vital capacity VC is greatly decreased in both stiff and compliant lungs.

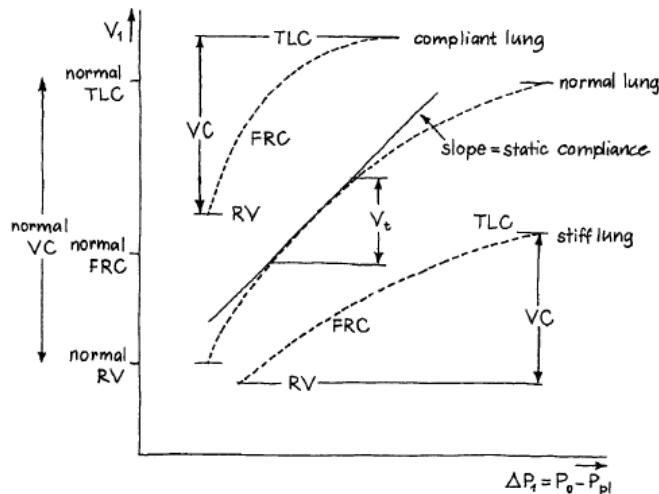


Figure 6-12 Static (no flow) pressure-volume curves of normal, stiff and compliant lungs.

The lung compliance can also be measured during breathing. When the patient breathes very slowly, the pressure drop over the airways is small and the alveolar pressure is almost zero, if the glottis is kept open. The resulting pressure-volume curve is the narrow loop of *Figure 6-13*. When a line is drawn through the points where the flow is zero (the end-expiratory and end-inspiratory points), its slope accurately reflects the static lung compliance.

When the patient breathes more rapidly, the PV-loop opens and becomes wider. The compliance computed from such a curve is called the *dynamic lung compliance*. Only for very slow breathing is it equal to the static lung compliance. The dynamic compliance is usually estimated by connecting a line between the two points on the loop (points P,Q) where the flow is zero. Please note that the dynamic compliance is not a real compliance, it is a clinical parameter that has found use during patient monitoring and estimation of the work of breathing.

The curves show hysteresis which becomes larger when breathing volume and flow rates increase. The hysteresis is related to energy dissipation processes. Studies have shown that the most important parameters related to the hysteresis are the visco-elastic properties of the lung tissue, friction in the intra-pleural space during lung during breathing and the airway resistance. Compliance estimated from P-V loops during continuous breathing is not accurate, the impact of resistive processes is too large. During mechanical ventilation special maneuvers have been developed to estimate lung compliance during no-flow static conditions. These will be discussed in module 5.

If no intra-esophageal pressure is available, a *dynamic lung-thorax compliance* can be computed in spontaneously breathing or artificially ventilated patients. *Figure 6-12* showed that the lung compliance decreases at large lung volumes. Lung over distension occurs when the lung is forcefully filled (e.g. by a ventilator) to an excessive volume, where its compliance is small. This is visible in the PY-loop, which then has the shape of a duck bill (*Figure 6-14*). Such a curve signifies that little volume increases result in large pressure increases, which could cause barotrauma. A duck bill PV-loop

indicates that the inspired tidal volume should be decreased; if the minute volume is to remain the same, the respiration frequency can be increased.

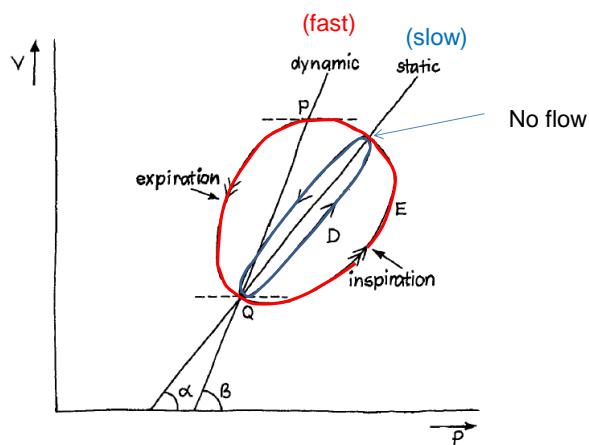


Figure 6-13 Dynamic pressure-volume loop, from which the lung's dynamic compliance can be determined. The latter varies with respiration rate.

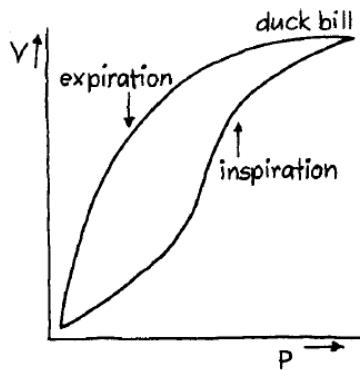


Figure 6-14 Lung over distention results in a "duck bill" PY-loop.

Although the PV-loop is generally useful, and is used by many clinicians during mechanical ventilation, a dynamic lung compliance computed from it may be unreliable. This is especially relevant for patients with obstructive and restrictive lung diseases. In both cases air passages have a larger (local) resistance; a model with a single resistance and a single compliance is now inadequate, since it assumes a uniform lung. An example will demonstrate this. Assume that the two lungs have equal compliances, but that the airway resistance of one lung is 10 times larger than that of the other lung (*Figure 6-15 (a)*). Yet, our model is that of *Figure 6-15 (b)*, which assumes *one* resistance R_{eq} and *one* compliance C_{eq} . At very low frequencies and flow rates, the pressure decrease due to the resistances is very small, and thus the dynamic compliance C_{eq} will be close to $2C$. At higher frequencies, the pressure drop across the resistances becomes larger. Because one lung has an RC time constant 10 times larger than the other one, it will take 10 times longer to fill it. If this time is not available due to the high respiration frequency only one lung is functional and C_{eq} will approach the value C . Thus, the value of the apparent lung compliance will vary between C and $2C$, depending on the breathing rate.

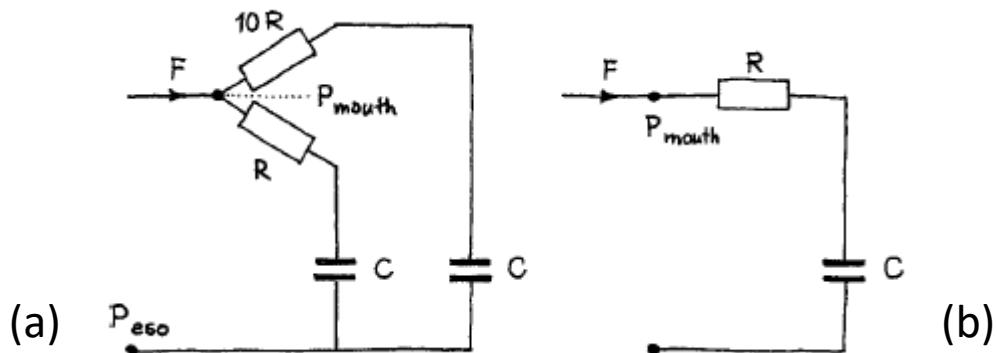


Figure 6-15 If the two lungs have different airway resistances and/or compliances, they must be modeled separately. The composite model on the right is now incorrect: there are two time-constants.

This variability of the apparent dynamic lung compliance can be usefully employed as a test for unequal airway resistances. A curve of the apparent compliance, usually measured at several respiration frequencies between 10 and 60 per minute, will provide information about unequal constrictions.

6.7 Pulmonary Function Tests

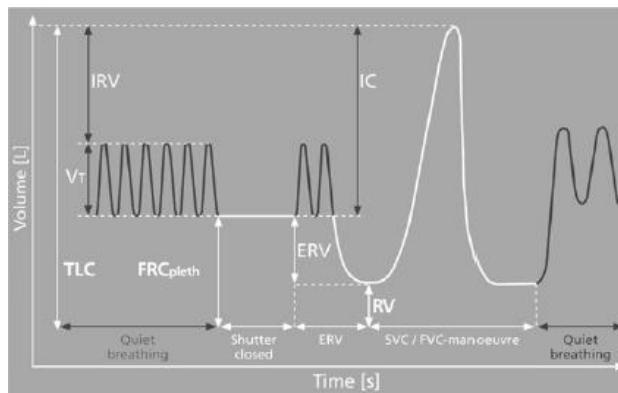
Pulmonary function tests provide information on the status and pathologies of the lung and airways. The airway resistance is very important parameter of the respiratory system and is of major importance for the pulmonary function. The airway resistance depends strongly on lung volume, can be different during inspiration and expiration and is affected by lung pathologies. The measurement of the airway resistance during slow and shallow breathing is described first. Thereafter pulmonary function tests which relate maximum air flow and lung volume are described.

6.7.1 Airway resistance

The airway resistance is defined as the ratio of the driving pressure difference (difference between the airway opening and alveolar space) and the rate of airway flow. The measurement of the airway resistance is based on Ohm's law.

$$R_{aw} = \frac{P_{awo} - P_a}{\dot{Q}} \quad 6.7-2$$

Measuring air flow rates \dot{Q} and the pressure at the mouth or airway opening P_{awo} is straightforward with the pressure and flow sensors that were described in module 3. The pressure measurement in the alveolar region P_a is more difficult but it can be estimated from measurements in the body-plethysmograph (see Figure 6-16). There are several options to measure the airway resistance, the method using the whole body plethysmograph is discussed below.



shutter closed

shutter open

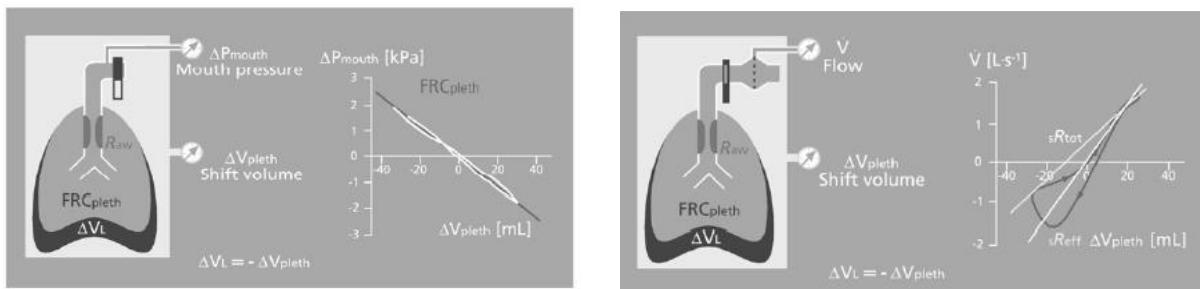


Figure 6-16 Diagram of a measurement of the airway resistance using a whole body plethysmograph.

Initially the test person breaths quietly from the free residual capacity FRC with small tidal volume V_T and normal rates. The FRC was determined previously by the closed shutter procedure described in the section on whole body plethysmography. During the closed shutter procedure the pressure at the mouth equals the alveolar lung pressure as there is no flow. Due to the inspiratory muscle effort the lung volume increases slightly and the lung pressure decreases. This is due to the decompression of air. There is no flow and the lung and airway pressures are proportional to the pressure generated

by the inspiratory muscles. Since there is no air flow Boyle's Law can be used to estimate the small change in lung volume. For example for an FRC of 2.5 liter and a lung pressure change of -10 cm H₂O (i.e. from an absolute pressure of 1000 cmH₂O to 990 cmH₂O) the resulting volume change is only 25.3 milliliter. This small change is called shift volume, it is a very important parameter that is also relevant during tidal breathing. Due to the increase in lung volume the pressure in the box rises with an amount ΔP_p and using Boyle's law the corresponding box shift volume change ΔV can be computed (it is of equal magnitude but of opposite sign as the change in lung volume). A plot of the lung/mouth pressure versus the measured shift volume is shown in the figure ((see Figure 6-16, shutter closed). This is an important result and links the lung pressure to the shift volume.

Then the shutter is opened and the test person is asked to breath quietly from FRC for a few breaths. Since air is inspired from the box its air volume decreases but at the same time the chest expands. However, the inspired air volume is slightly smaller than the increase in lung volume due to tidal breathing. The difference is equal to the shift volume and it measured directly by measuring changes in box pressure and calculating the shift volume using Boyle's Law. The flow rate is plotted versus the measured shift volume ΔV ((see Figure 6-16, shutter open). A flow-shift volume loop is observed. From this loop the so-called "specific airway resistance sR_{aw} " can be obtained from the inverse slope of the flow-shift volume curve. This parameter is equal to:

$$sR_{aw} = P_{atm} \cdot \frac{\Delta V_{box}}{\dot{V}} \quad 4.7$$

It is not real resistance but it is an important parameter for lung (patho)physiology. This parameter can be determined in several manners from this loop, each having its own merits and use cases (see Figure 6-16).

The shift volume can be converted to lung pressure by using the results from closed shutter procedure. Using Boyle's law and some rearrangement this gives the following result:

$$R_{aw} = \frac{sR_{aw}}{FRC} \quad 4.8$$

The airway resistance can simply be obtained by dividing the specific airway resistance by the free residual capacity. A typical value for the airway resistance for a healthy male adult is between 1 and 2 cmH₂O.s/L. For test persons with obstructive and restrictive lung disease the resistance can be either much higher and it can depend on inspiration or expiration (COPD, lung emphysema, asthma) or it can be lower (lung fibrosis). Furthermore the shape of the flow-shift volume loops provides important information to the clinician needed for diagnosis, therapy, therapy adjustment and evaluation of medicine treatments.

Flow-shift volume loops are plotted for four subjects (see Figure 6-17 (I)). For the normal subject (1) the inspiration and expiration traces are almost identical, the airway resistance during inspiration and expiration is roughly equal. For subjects with lung diseases the loop shape differs markedly. From these loops several parameters can be extracted. It is possible to extract the airway resistance as function of lung volume. A measurement of the airway resistance during inspiration and expiration of subject (3) with chronic airway obstruction (COPD) is shown in Figure 6-17 (II). The airway resistance is much higher than that of the reference group during both inspiration and increases strongly during expiration. During expiration the airway resistance increases due to collapse of the medium sized airways of this subject. This is a very complex phenomenon. The collapse is related to two effects. The first is the reduction in elastic pull/stretch of the connective tissue in the COPD lung which reduces the diameter of these airways. This effect in combination with the change in

intrapleural pressure and large “Ohmic” pressure drop in the airway during active expiration effort causes a local collapse of the airways and leads to a very large expiratory airway resistance. This will be described in more detail in the following section on maximum effort lung tests.

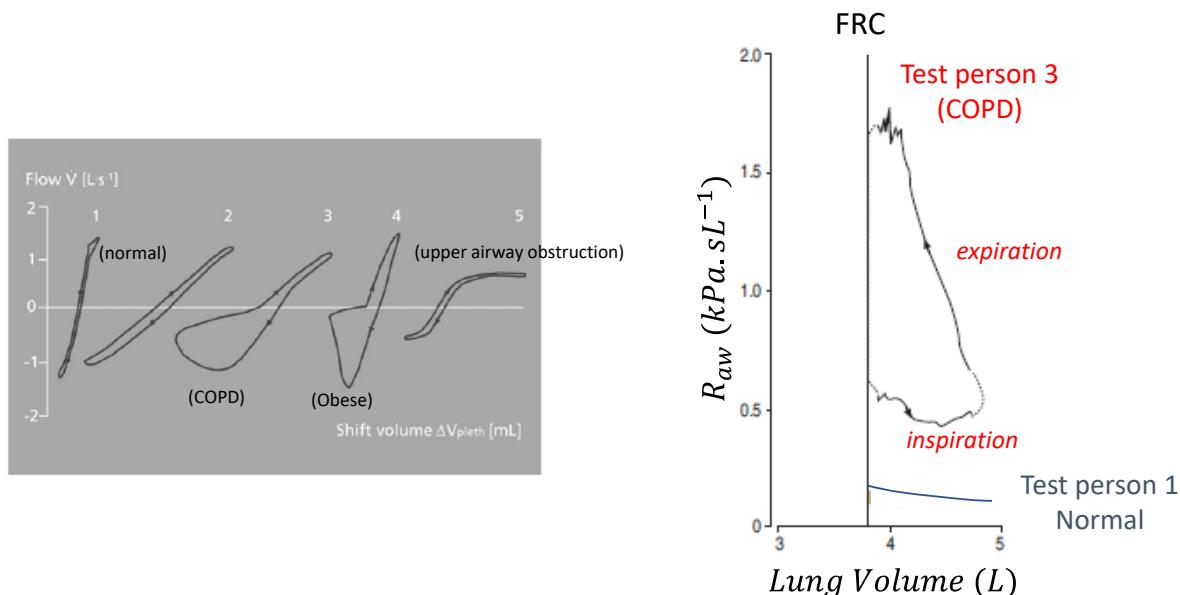


Figure 6-17 (I) Flow-shift volume loops of four subjects: (1) normal subject, (2) Subject with chronic high upper airway resistance, (3) Subject with chronic airway obstruction, (4) high BMI obese subject and (5) subject with upper airway obstruction. (II) For subject (3) the airway resistance during inspiration and expiration is shown.

6.7.2 Forced Expiration Test

In Figure 6-18 various inspiration and expiration maneuvers are shown which can be measured with a spirometer. Forced inspiration and expiration tests with a spirometer are simple but very useful tests. The patient is asked to either inhale or exhale to a maximum or minimum lung volume. Thereafter a maximum expiration or inspiration effort test is done. The air flow rate and change in lung volume are recorded simultaneously.

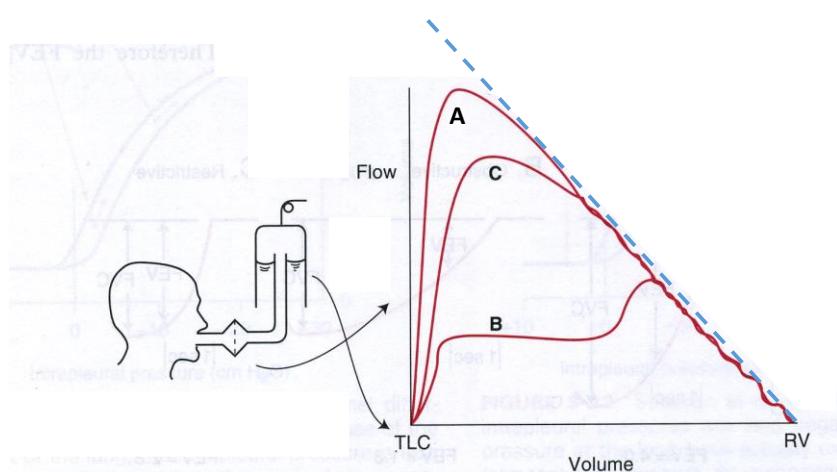


Figure 6-18 Flow rate versus lung volume for (A) maximum inspiration followed by maximum expiration, (B) Initially slow expiration followed by a maximum expiratory effort and (C) a second maximum expiratory effort directly after effort (A).

A plot of the air flow rate versus expired air volume is shown in Figure 6-18 for a maximum effort expiration test. Three measurements are shown for the same subject. In curve (A) the flow was

measured after a maximum inspiration to the total lung capacity followed by a maximum expiration effort. The flow increases initially with a fast rate, reaches a maximum and decreases until the residual lung volume is reached. Directly after the first test, the test is repeated, this gives curve C. Due to fatigue (reduced maximum muscle effort) the peak flow is lower, but the two curves are identical for smaller lung volumes. Finally, the subject is asked to reduce expiration effort and to exhale with maximum effort later on. This gives curve B. Note that the region right of the blue dashed line in the airflow-lung volume plot area cannot be reached by this test person whatever the expiratory effort that is applied. This line is specific for the test person. There is a physical effect that limits maximum flow rate at given lung volume, at a given lung volume the maximum flow rate does not depend on effort. The physiology is related to the flow in elastic collapsible tubes when the tube pressure is larger than the external pressure at the inlet side and smaller than the external pressure at the outlet side (see section 4.6).

The origin of the flow limitation and effort independence during expiration is related to the dynamic compression and collapse of the elastic small and intermediate airways by the intrathoracic (i.e. intrapleural) external pressure generated by the muscles. Data of the dependence of airway flow of intrapleural pressure at fixed lung volume confirms this hypothesis. Note that the intrapleural pressure is proportional to the muscle effort of the test persons. The intrapleural pressure can be measured with an esophageal balloon catheter. The flow, lung volume and intrapleural pressure are measured during a series of tests. From this data set three isovolume curves for expiratory and inspiratory flow versus esophageal pressure (proportional to muscle effort) were constructed from a large series of measurements, the results for three fixed lung volumes are plotted in Figure 6-19. The lung volume is a parameter (small, normal, large).

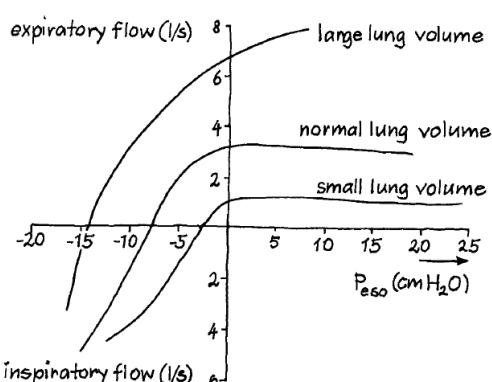


Figure 6-19 Expiratory flow as a function of lung transmural pressure at three different fixed lung volumes. At small and normal lung volumes, expiratory flow is effort independent.

For a given very large lung volume the expiratory flow measured at isovolumic conditions increases with effort, the expired flow remains effort dependent. For smaller lung volumes, i.e. for a normal and small lung volume the expiratory flow reaches a plateau and becomes effort independent. Note that the inspiratory flow is effort dependent for all lung volumes. A first order explanation of the maximum flow rate at a given fixed lung volume is illustrated schematically in Figure 6-20 for the normal lung volume.

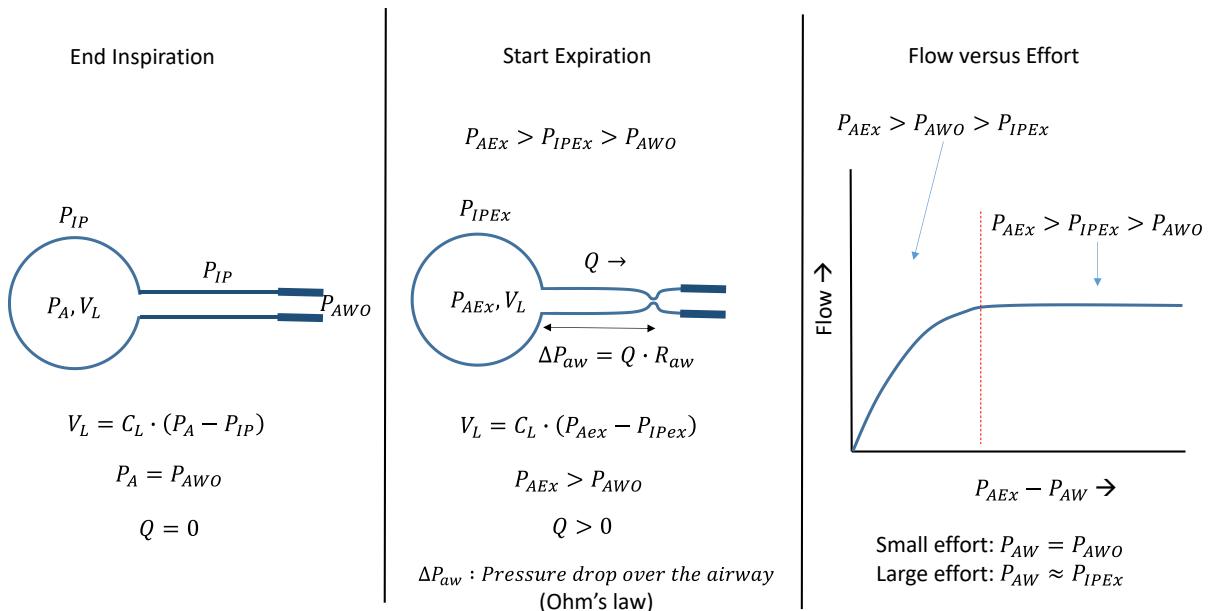


Figure 6-20 Diagram to illustrate the dependence of expiratory airflow on effort at isovolumic conditions. The thick line in the diagrams represents the part of the larger airways that is covered with cartilage rings. This part of the airway cannot collapse.

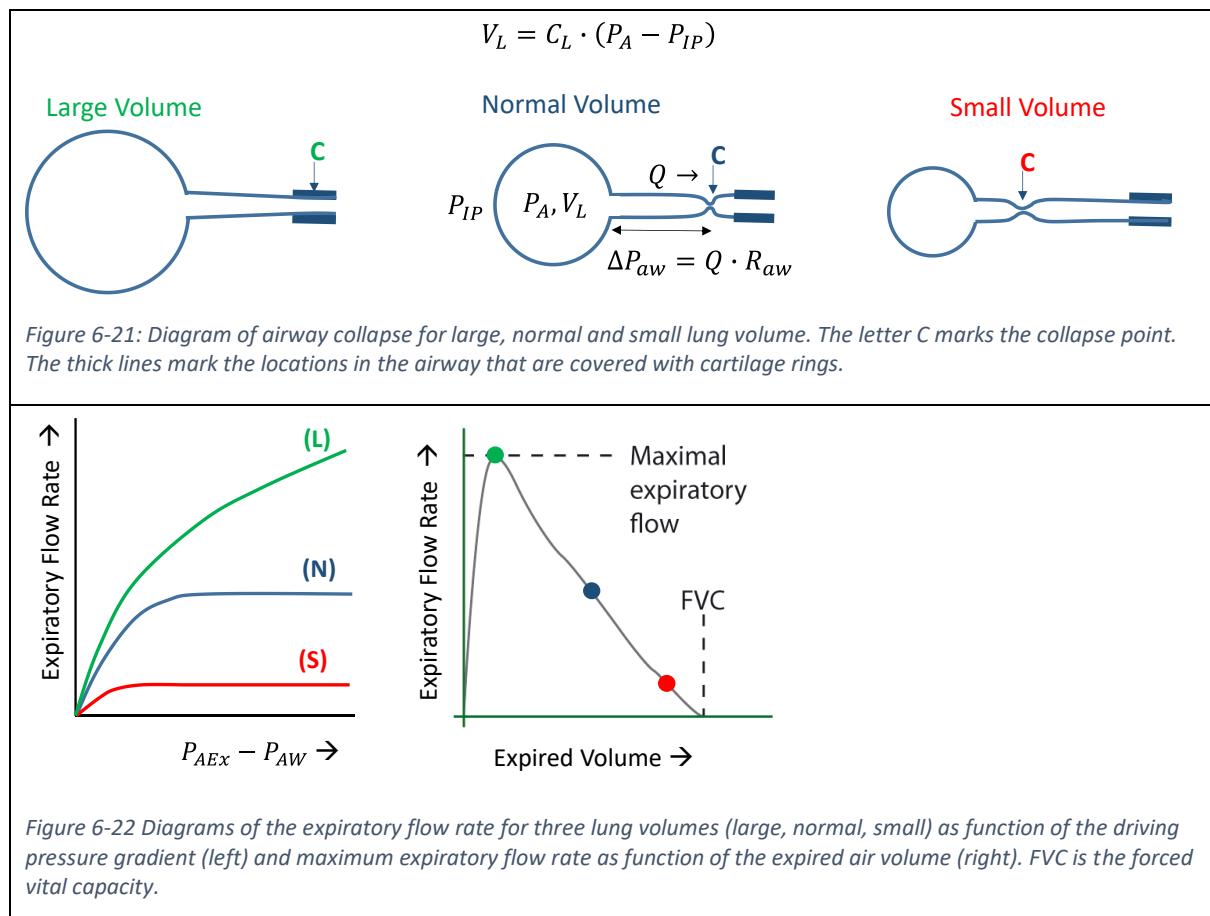
At the end of inspiration, the inward flow is zero and the final lung volume has been reached, this volume is equal to the product of lung compliance and the transmural pressure (i.e. the difference between alveolar and intrapleural pressure). Next the effort dependence of the expiration phase is described under isovolumic conditions.

The lung expiration starts at the conditions that were reached at the end of the inspiration described above. Due to the expiratory effort initially the intrapleural pressure and alveolar pressures both increase with the same amount¹¹. Note that when there is no flow directly after the change in muscle effort, the lung volume has not changed and the transmural pressure does not change, i.e. the intrapleural pressure and alveolar pressure increase with the same amount. Since alveolar pressure is larger than the AWO pressure, air starts to flow towards the airway opening (AWO). Note that the lung resistance is a distributed resistance, in first order it can be approximated by the sum of many small resistances in series. From Ohm's law it follows that when there is flow, a pressure gradient must be present over the airway resistances, the pressure drop depends on the local airway resistance. The airway pressure decreases over the airway towards the AWO where it is equal to atmospheric pressure. Note that the intrapleural pressure is in first order¹² constant over the lung and elastic airway regions, they are within the pleural space region. Hence the transmural pressure of the airway decreases in the direction of the AWO. When the muscle effort is sufficiently large at some point in the airways the local transmural pressure becomes zero. At this point the tubes of the airway will collapse, the local resistance is very large. When collapse occurs in the airways the driving pressure for flow is equal to the difference of the alveolar and intrapleural pressures. Although both pressures depend on the magnitude of the effort and lung volume, for isovolumic conditions the difference between the two pressures is constant and as a result the flow rate does not depend on the magnitude of the effort as the driving pressure gradient does not depend on the effort anymore. This is illustrated in the right part of Figure 6-20. This phenomenon is equivalent with the waterfall

¹¹ There has been no flow and lung volume and transmural pressure have not changed yet!

¹² We exclude effects of gravity on the intrapleural pressure

effect in elastic tube flow when external pressure is in between the inlet and outlet pressures (see 4.6). Next the dependence of maximum expiratory flow rate on lung volume is discussed.



The maximum expiratory flow rate at two isovolumic cases is discussed, i.e. a smaller and much larger lung volume. The normal lung volume was discussed before. First consider a smaller lung volume. A smaller lung volume corresponds with a smaller transmural pressure over the lung. When expiration starts, air flows toward the AWO and this flow is accompanied by a pressure drop over the airways. As discussed for the normal case, collapse in the airways occurs when the airway pressure at a certain point is equal to the intrapleural pressure. Note that a smaller pressure-drop over the airways is needed for airway collapse to occur in the small lung volume case (i.e. the difference between alveolar and intrapleural pressure is smaller than for the normal lung). Hence the collapse point shifts to a location deeper in the lungs (see Figure 6-21) and the driving pressure for flow is smaller than for the normal lung case. Therefore, for the isovolumic small lung volume case the maximum expiratory flow is smaller than for the normal lung (see Figure 6-22). This can be generalized for other lung volumes, when collapse occurs in the airway during a maximum expiratory effort the maximum flow rate depends on lung volume.

Next consider the case of a very large lung volume. For this case the transmural pressure over the lung is much larger than for the other cases and the point where collapse can occur shifts towards the airway opening, it can be even outside the thoracic cavity. The point C where the airway pressure is equal to the intrapleural pressure occurs in the larger airways. At this site cartilage rings surround the airways and airway collapse is prevented (see Figure 6-21). Therefore, the expiratory flow at fixed large lung volumes is large and remains dependent on the effort (see left part of Figure 6-22).

A diagram showing the maximum expiratory flow versus expired air volume is shown in the right part of Figure 6-22. During expiration the lung volume decreases and as discussed above the maximum expiratory flow will decrease as well. This is illustrated for the three lung volumes discussed above. The symbols in this figure correspond with the maximum flow rates that can be reached for the three lung volumes. An increase in expiratory effort will not lead to an increase in flow due to airway collapse (waterfall effect). For very large lung volumes collapse does not occur and expiratory flow remains effort dependent. This model qualitatively explains the expiratory flow-pressure curves in Figure 6-18, during expiration the lung volume decreases and the corresponding maximum flow will decrease as well. For small expiratory effort collapse does not occur in the collapsible part of the airway and expiratory flow rate will depend on expiratory effort (curve B). The differences between curves A and C at small expired air volume (i.e. large lung volume) are caused by fatigue of the expiratory muscles, the maximum effort during test C was smaller than that for the first test A. For very small lung volumes close to the residual volume RV, the chest wall is very stiff which limits further volume reduction. Furthermore, at the RV volume, the airway collapse occurs almost immediately after the expiratory effort begins, airflow is blocked. Both chest wall elastic properties and airway collapse limit reduction of the lung volume below RV. From the above it is clear that maximum effort expiration curves depend on factors like lung volume, lung and chest wall compliance, airway resistance and collapse of the airways. These simple measurements are very useful to study the lung performance and lung pathologies. This is described in the next section.

6.7.3 Test of lung pathologies

The forced expiration test described in the previous section is very useful test of the pulmonary function that is used routinely in clinical practice. These tests show clear differences between different lung pathologies such as obstructive and restrictive ones. Important parameters are the maximum forced expiratory volume in one second (FEV1) and forced vital capacity (FVC). After inspiring to the maximum lung volume, people are asked to exhale air with maximum effort. Qualitative example plots are shown in Figure 6-23 for subjects with normal pulmonary function and for subjects with restrictive and obstructive lung diseases. Typical values for FEV and FVC are indicated for the three cases. Two common deviations in the patterns can be observed. For the restrictive diseases such as pulmonary fibrosis both FEV and FVC are reduced, the time to reach these values is comparable to the normal case. The ratio of FEV to FVC is normal or increased. In obstructive diseases such as asthma FEV1 is reduced much more than FVC, the ratio FEV1/FVC is smaller than for the normal case. The FVC is reduced because of the trapping of air due to early airway collapse. The expiration time is increased due to the increase in airway resistance during expiration.

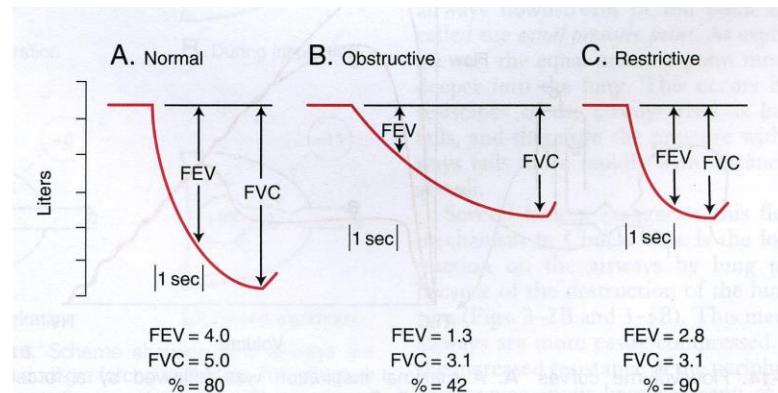


Figure 6-23 Typical expired volume versus time plots for subjects with normal lung function and with lung function degraded by obstructive or restrictive lung diseases.

These measurements become even more useful when combined with whole body plethysmography. Plots of flow versus absolute lung volume are then available (see Figure 6-24).

In restrictive lung diseases the residual volume is smaller than for the normal case, peak flow is reduced because the lung volume is smaller. The flows at small lung volumes can be abnormally high due to the recoil of the stiff lungs and decreased airway resistance at small lung volumes compared to the normal lung. For the obstructive lung diseases flow rates are much lower than in the normal case. They are very small given the large lung volumes. In restrictive cases the lung volume is limited by the stiff lungs, chest wall or weak inspiratory muscles. For the obstructive cases the total lung capacity is abnormally large, but the vital capacity is reduced by the early collapse of the smaller airways. The option to measure lung resistance in combination with expiratory respiration maneuvers is another advantage of the whole-body plethysmograph, it has however not found widespread use yet.

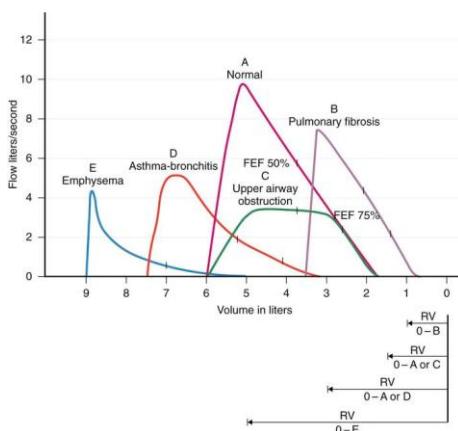


Figure 6-24 Volume-flow curves for maximum effort expiration after maximal inspiration.

6.8 Diffusion Capacity

The lung's O₂ diffusion capacity DO₂ is defined as the O₂ flow through the lung membrane towards the alveolar capillaries divided by the O₂ partial pressure difference across the alveolar membrane. The average O₂ flow through the lung membrane is the patient's O₂ consumption, which can be measured by spirometry. To establish the O₂ partial pressure difference across the lung membrane, we would need to measure the O₂ partial pressure in the arterial gas and the O₂ partial pressure in the blood that perfused the alveoli. The former can be determined from the end expiratory O₂ concentration, the latter in principle from a blood sample. Since blood samples can be obtained from large blood vessels only, it would not be completely clear whether the arterial or the venous PO₂ should be used or an average of both.

The following non-invasive method is used clinically. It is known that carbon monoxide (CO) has almost the same diffusion properties as O₂ when it passes the alveolar membrane, i.e. D_{CO} ≈ DO₂. We also know that CO, once in the blood, is rapidly bound to hemoglobin (210 times faster than O₂); its partial pressure in the blood is therefore effectively zero. For the diffusion capacity carbon monoxide is used instead of oxygen, but since CO is toxic, very low concentrations must be used. The patient inspires a gas mixture which contains traces of CO (0.3%) and He (10%); the inspiration is from RV to TLC. After 10 seconds, the patient exhales again to RV. He is used for calibration purposes. Because of Helium's very small diffusion capacity, we expect the inspired He to be completely exhaled again.

We also expect the CO to have partially disappeared. During the 10 seconds, the alveolar CO concentration must have decreased exponentially (see Figure 6-25).

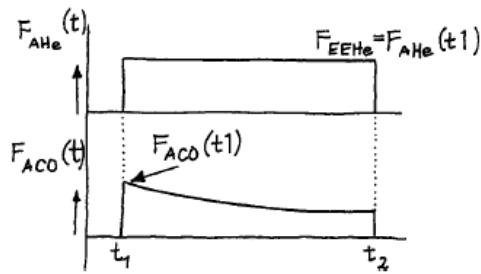


Figure 6-25 CO diffuses rapidly through the alveolar membrane, so that its alveolar concentration decreases. He cannot pass the lung membrane, so its alveolar concentration remains unchanged.

The diffusion capacity for carbon monoxide is obtained from the equation:

$$Q_{CO}(t) = V_A \frac{df_{CO_A}}{dt} = D_{LCO} \cdot f_{CO_A} \cdot (P_{atm} - P_{water}) \quad 6.8-1$$

The solution for the alveolar fraction of CO is:

$$f_{CO_A}(t) = f_{CO_A}(t_1) \cdot e^{-D_{LCO} \cdot (P_{atm} - P_{water}) \cdot (t - t_1) / V_A} \quad 6.8-2$$

The diffusion capacity can be obtained from this solution at time t_2 by using the helium fractions in the inspired and expired air which gives:

$$\frac{f_{CO_A}(t_1)}{f_{ICO}} = \frac{f_{He_A}(t_1)}{f_{IHe}} = \frac{f_{He_A}(t_2)}{f_{IHe}} \quad 6.8-3$$

Here f_{ICO} and f_{IHe} are the known fractions of CO and He in the inspired gas mixture (0.3 % and 10%). The alveolar volume can be obtained by the helium dilution method.

Note that this method expresses the diffusion properties of the alveolar membrane as a single number. This number cannot give insight into any possibly existing local differences.

6.9 Blood gases and pH (Optional)

Blood gases are a group of laboratory tests that are done on arterial and mixed venous blood samples. Partial pressures of oxygen, carbon dioxide and pH of blood are measured using electrochemical techniques. Furthermore, concentrations of bicarbonate ions, lactate ions, hematocrit, O₂-saturation of Hemoglobin and concentrations of ions (Na, K, Ca, Mg) are easily measured. In clinical practice either syringes or multi lumen catheters are used to obtain blood samples. For arterial blood the radial artery is mostly used, it is the most practical location. For mixed venous blood the Schwan-Ganz catheter is used, blood samples are taken from the pulmonary artery. Blood gases are a very useful test to investigate the function of both the respiratory and cardiovascular systems. The drawback is invasiveness and obtrusiveness of the method. Therefore, these tests are done mainly in high acuity care settings (surgery, emergency care, intensive care).

6.9.1 Capnography and Pulse Oximetry

Capnography and pulse oximetry provide important and continuous information on important clinical parameters. The ease of use and quality of information are attractive features. These very useful monitoring techniques are routinely used during a wide variety of care settings and have proven to be of great value during surgery and critical care. Pulse oximetry can also be done in home care environments. Pulse oximetry gives information on pulsatile blood flow, the pulse rate of the heart and on the saturation of hemoglobin. Capnography provides information on the partial pressure of carbon dioxide in arterial blood, the respiration rate and status of the respiratory and circulatory systems. It is widely used to control artificial ventilation of patients. In Figure 6-26 an image of a Philips IntelliVue X3 monitor is shown with the waveforms and numeric data of ECG (green), SpO₂ (blue), capnography (white) and numerical values of non-invasive blood pressure and rectal temperature. Clinicians are interested in both the waveforms and numerical values like SpO₂ and ETCO₂. The waveforms are important because they contain a lot of analog information and also are an indication of the quality of the measurement.

A pulse oximeter is most likely the first device that is applied to a patient during surgery or emergency care. Within a few seconds the clinician has information on the pulse rate, that the heart pumps blood into the arterial system and of the oxygenation of arterial blood. Capnography is more difficult and more invasive and requires intubation. It provides extremely useful real time data to the clinician. The routine use of capnography during surgery has been a main factor for the strong reduction of mortality during and after surgery.



Figure 6-26 Philips IntelliVue X3 monitor with waveforms and numerics of ECG (green), pleth waves (blue), capnography (white) and numerical values of noninvasive blood pressure and temperature.

6.9.2 Respiration Rate

The respiration rate of non-ventilated patients is of great importance for clinicians. Respiration rate can be an early sign of patient deterioration preceding severe illness. Either very low (< 8 per minute) or very high rates (> 25 per minute) are relevant. The former for respiratory depression which can be fatal, the latter is an early indicator for sepsis, pneumonia, lung embolism and hemodynamic instability that are life threatening conditions. Note that abnormal respiratory rates values are a warning for imminent patient harm, in many cases there are other causes that might be less harmful.

For a high-acuity patient the ventilation rate is determined by the mechanical ventilation machine and is known. For non-ventilated patients the respiration rate is difficult to measure manually, there are several monitoring options. The most frequently used measurement is bio-impedance which is

combined with the ECG measurement (chapter 5.7). This measurement is prone to motion artifacts and clinicians have developed an antipathy for this measurement especially for mobile patients. A second and more reliable method is side stream capnography, but this method is not well suited for mobile patients. For sleep studies Respibands and air flow sensing near the nostrils are used. These sensors are not well suited for mobile patients.

Recently there has been a growing interest to measure respiration rate of patients in the hospital ward. The goal is early detection of patient deterioration in a low acuity environment as the hospital ward or home. New measurements that use accelerometers to extract heart rate and respiration rate have been introduced recently. These sensors are used in a cable-less measurement from Philips for monitoring of low acuity patients in the ward. The device is shown in Figure 6-27. A small box with a 3-axis accelerometer, micro controller, wireless connectivity and a battery is applied at the chest. The optimum location is on the left part of the chest (see Figure 6-27) just below the ribs. The chest moves due to the respiration and blood pressure pulses and the chest motion causes changes of the accelerometer position with respect to the earth gravitational field. These changes in gravitation acceleration components are measured and heart rate and respiration rate can be extracted. This sensor is useful when the patient is at rest, during motion the data is not valid. The data during motion period are rejected and respiration rate can be measured regularly and with good accuracy and precision when compared to the reference device. The performance is as good or better as bio-impedance and acoustic measurements. The sensor can also be used for posture, activity and fall detection measurements. Recently an improved version has been developed and is available in the Benelux area (Philips Healthdot). This device has a battery life longer than three weeks and does not require a hub for the wireless connectivity.

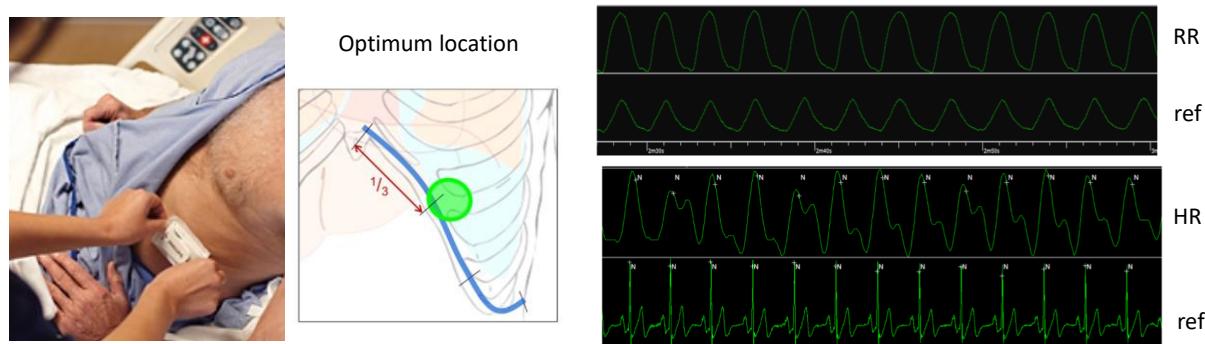


Figure 6-27 Respiration and heart rate sensor from Philips Cable-less measurements. Left application to a patient, mid optimum location, right signals and reference signals for respiration and heart rate.

A second option for low acuity and home measurements of the respiration rate is a small patch with a short lead ECG, bio-impedance and accelerometer. The combination of ECG, bio-impedance and accelerometry gives the option for sensor fusion and more robust parameter extraction. An example from Philips is shown in Figure 6-28. The sensor is designed for use duration of multiple, typical is an average length of stay in the hospital. The requirement is battery life large than three days. The material cost is however significantly higher than for the accelerometer option. Furthermore the ECG electrodes pose problems with skin irritation when longer use time is needed.

Finally, there is a large activity to extract the respiration rate from the modulation of the optical pleth signal. This would also enable measurement of the respiration rate using watches on the arm. Although encouraging results have been reported it is not clear if accuracy and precision can be guaranteed for conditions as typically occur during low acuity monitoring.

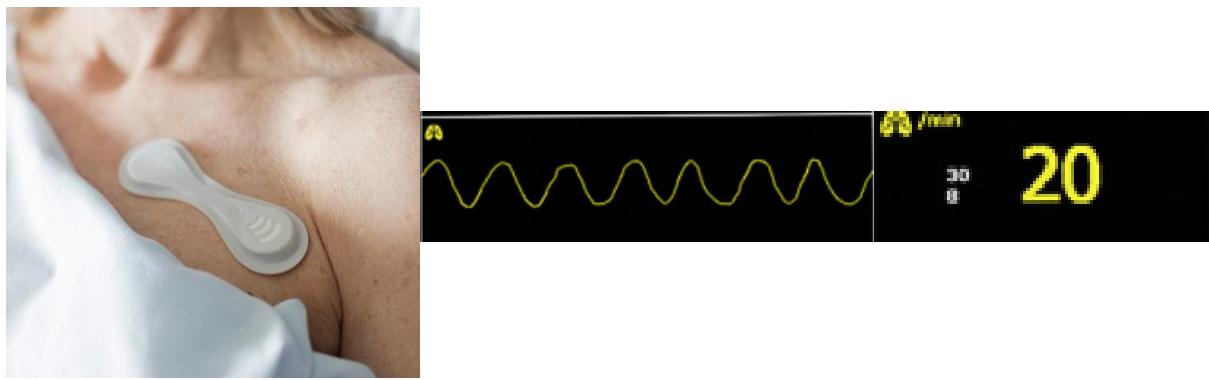
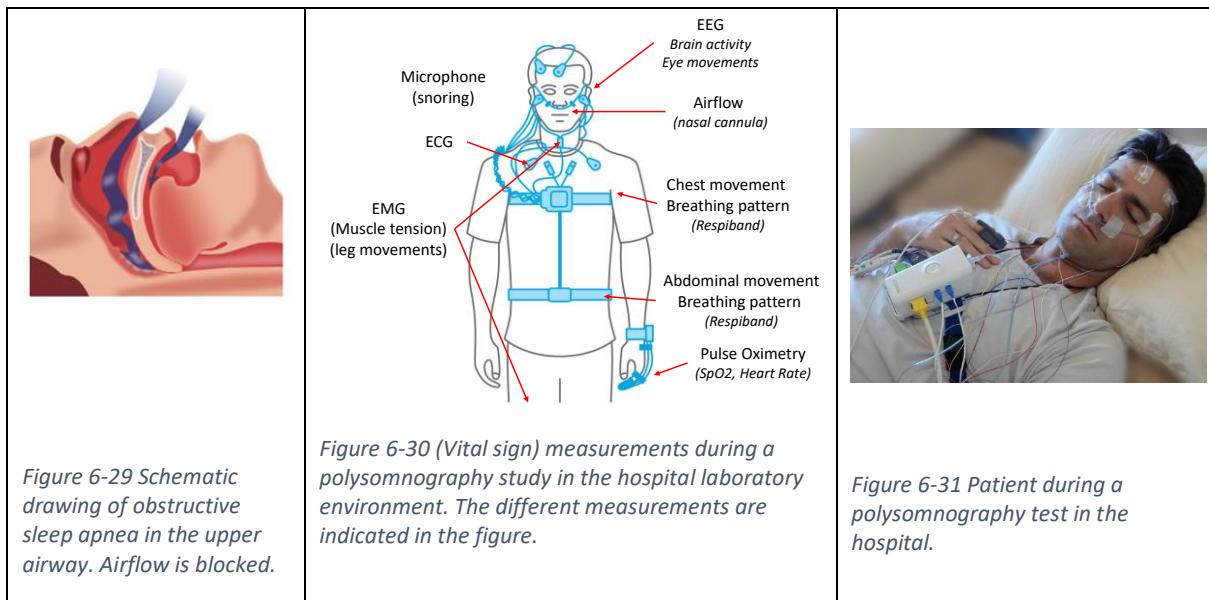


Figure 6-28 Philips biosensor for low acuity measurement of heart rate and respiration rate. Left location on the chest, mid the signal and right the extracted respiration rate.

6.10 Measurements During Sleep Disorders

It has been estimated that approximately 25% of adults worldwide suffer from (obstructive) sleep apnea (OSA). About 80% are underdiagnosed. (O)SA is a comorbidity to most cardio-vascular diseases. During sleep apnea breathing regularly stops and starts during sleep. These pauses can last for more than 10 seconds. In obstructive sleep apnea the throat muscles intermittently relax and the airway in the throat is blocked (see Figure 6-29). OSA is a serious sleep disorder. Note that for OSA there is still an effort from the respiratory muscles, but airflow is blocked.



For diagnosis of sleep apnea and the type of apnea many physiological parameters must be measured and subsequently analyzed and interpreted by a clinician. This is mostly done in a hospital environment during the night in a sleeping period. Such a study is called polysomnography. The measurement includes measurement of the breathing effort, rate and patterns using Respibands on the chest and abdomen, airflow in the nasal area either by a nasal cannula or a thermistor (changes in air temperature) and pulse oximetry mostly at the fingertip (blood oxygenation SpO₂, heart rate). This gives insight in the respiratory parameters and type. When a nasal cannula is used it is possible to use side stream capnography and ETCO₂ can be obtained. From the capnogram the respiration rate can be easily determined. During respiratory depression or obstruction, the ventilation of the lung is absent and SpO₂ will drop and when lung ventilation is resumed the ETCO₂ value will be increased. ECG is also frequently measured to study a cardiovascular disease that could cause or be

related to the sleep apnea. To monitor sleep quality the activity of the brain (EEG), eye movement (EEG) and muscle relaxation at the chin (EMG) is measured.

Note that the polysomnography test is quite obtrusive for the patient, there are many measurements and leads attached to the patient. There is a large research activity to use smart devices such that at least screening of sleep apneas can be done in the home environment.

6.11 References

[West] John B. West, Respiratory Physiology, The Essentials, Wolters Kluwer, eight edition, ISBN 9780781772068.

[Blom] J.A. Blom, Monitoring of Respiration and Circulation, CRC Press LLC, ISBN 9780849320835

6.12 Questions

1. What is an obstructive lung disease?
2. Asthma and COPD are both obstructive lung diseases, what are the main differences?
3. What is a restrictive lung disease?
4. Draw a schematic diagram of a spirometer and explain the function of each component.
5. Describe the main difference in flow-volume plots measured during a forced expiration test for patients with COPD, Asthma, normal reference subjects and patients with restrictive lung diseases.
6. Explain how a patient's oxygen consumption is measured with a spirometer.
7. The text describes three methods to measure residual volume. Which methods? Explain one of those methods in detail.
8. What is the best method to measure residual volume? Explain why.
9. Which measurements can be done with a whole body plethysmograph?
10. How is the airway resistance measured?
11. How is the dead space measured?
12. Explain the shape of a typical capnogram.
13. Describe the single breath nitrogen washout method.
14. What is the difference between the static and the dynamic lung compliance? How do the numeric values compare to the true lung compliance?
15. Explain why the MEFV curve has an effort independent part.
16. How is the lung's diffusion capacity measured?
17. Explain how a change in thorax compliance will likely change the lung's vital capacity.
18. Describe possible modifications of the spirometry methods described in this chapter so that they could be converted to monitoring methods for e.g. patients in an intensive care unit.
19. Compute the value of C_{eq} in Figure 6-15, and show that it is frequency dependent.
20. What are blood gases? Why are they so valuable to the clinician?
21. Which methods/sensors can be used to measure the respiration rate of non-ventilated and mobile patients? Describe one of the methods in detail.

7 Circulatory Measurements

The main function of the circulatory system is transport of oxygenated blood, nutrients and other substances towards the cells in the tissues and transport of deoxygenated blood to the lungs. During a disease these functions may be impaired and for the diagnosis and treatment clinicians need vital signs measurements and data related to the main functions of the circulatory system. In chapter the main measurements of the parameters of the cardio-vascular system (blood pressure, flow) are described.

7.1 Measurements on the cardio-vascular system

Measurements of the electrocardiogram, blood pressure and arterial blood flow rates are common in clinical practice. Measurements and interpretation of the ECG is called electrophysiology. It is a specialization of cardiologists and clinicians and falls outside the scope of this course. More information can be found in [Guyton] and [Webster]. In the following sections measurements of blood flow and blood pressure are described.

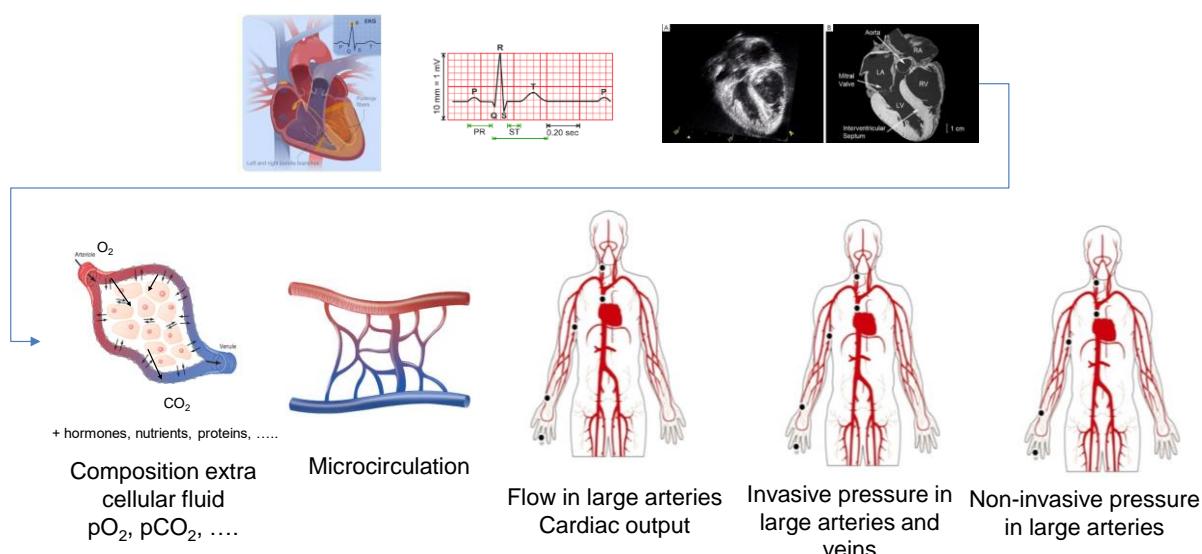


Figure 7-1 Diagram showing the order in the level of information versus the obtrusiveness to the patient.

In the diagram of Figure 7-1 the order of the clinical significance of parameters of the circulation is shown from the perspective of homeostasis. In many cases treatment is aimed at guaranteeing sufficient oxygen and nutrients delivery to the tissues. The in-vivo composition of the extra cellular fluid is the most important parameter to clinicians want to know. This parameter cannot be measured during clinical care. The next best option to estimate this composition is a measurement of blood flow in the micro circulation. This measurement can be done but is very complex and can be done only in rare cases in research studies. Capillary blood flow is a very important parameter for clinicians, but its measurement is unfortunately too complex, time consuming and is not compatible with workflow in clinical care settings. Therefore, other measurements are used in clinical practice.

The next option to estimate O₂ delivery to the tissues is measurement of the flow rate and blood composition in the large arteries of the systemic circulation. Measurement of the blood flow in the large arteries is highly invasive and poses high risk for the patient. It is measured only when the benefit for the patient outweigh the risks. This is done typically during complex and high-risk surgery and during care in the intensive care unit for selected patient groups.

The next level of information is blood pressure, it is a surrogate for flow. Continuous invasive arterial blood pressure would be most valuable but also poses risks for the patient and is again only used when the benefit outweighs the risk. Therefore, intermittent non-invasive blood pressure (NIBP) is the most common measurement. A typical measurement time interval is 5 to 15 minutes. Blood pressure measurements are used most frequently in clinical practice, and the techniques that are used in the clinic are described in the following section.

7.2 Blood Pressure

Blood pressure is defined as the force per unit area exerted on the wall of a blood vessel. It is the internal pressure exerted by blood on the viscoelastic vessel walls and is measured with respect to atmospheric pressure. The unit used in clinical practice is mmHg (133 Pa), the SI unit (Pa) is rarely used. A short description of the volume-pressure (V-P) relation of a blood vessel follows below. This V-P relationship is important for blood pressure measurements that are described later in this chapter.

7.2.1 Arterial Pressure-Volume Relation

In a static condition, blood pressure depends on the elastic properties of the blood vessel wall and on the blood volume in the vessel segment. A static volume-pressure relation and the differential compliance of a large artery are shown in Figure 7-2 (a). At a positive transmural pressure, the elastic arterial wall is stretched, the vessel is mechanically stable, its cross section is circular, its volume increases with increasing transmural pressure. The differential compliance is pressure dependent and decreases at higher pressures (the vessel becomes stiffer). It has a maximum compliance around zero transmural pressure. For decreasing transmural pressure, the shape remains circular, the radius decreases. Below 50 mmHg the vessel volume decreases strongly and at zero transmural pressure the unstressed volume has been reached, i.e. there is no mechanical wall stress the fluid volume is equal to the radius of the artery when there is no transmural pressure. The non-linear shape of the volume and compliance curves is related to the diameter, material wall thickness and composition (elastin, collagen, smooth muscle), smooth muscle tone and non-linear visco-elasticity of the arterial wall materials. For negative transmural pressure the mechanical behavior becomes very complicated, the vessel loses its mechanical stability, collapses and deforms in a complex manner. In the simplest case the shape becomes oval. At more negative pressure the midpoint touches and for even more negative pressures the central part collapses and two small lumens at the edges are formed. At very large negative pressures the vessel is completely collapsed. There is a variation of around 10 mmHg of the peak position (deviation from zero transmural pressure) from person to person. Note that this is important for oscillometric blood pressure measurement.

In Figure 7-2 (b) a magnification of a section of the V-P curve of Figure 7-2 (a) is shown. The black curve is the arterial pressure waveform with arterial pressure between diastolic and systolic blood pressure. The pressure in the tissue around the artery is atmospheric pressure, all pressure is dropped over the arterial wall. The variation in arterial pressure leads to a change in arterial volume (blue curve). The arterial volume waveform differs from the pressure waveform due to the non-linear volume-pressure relation. When an external counter pressure is applied to the blood vessel, the transmural pressure changes and the magnitude of the volume variations at each beat changes. This application of counter pressure and measuring the changes in arterial volume pulsations is exploited in most non-invasive blood pressure measurements. In the following section the main body locations where blood pressure is measured are described.

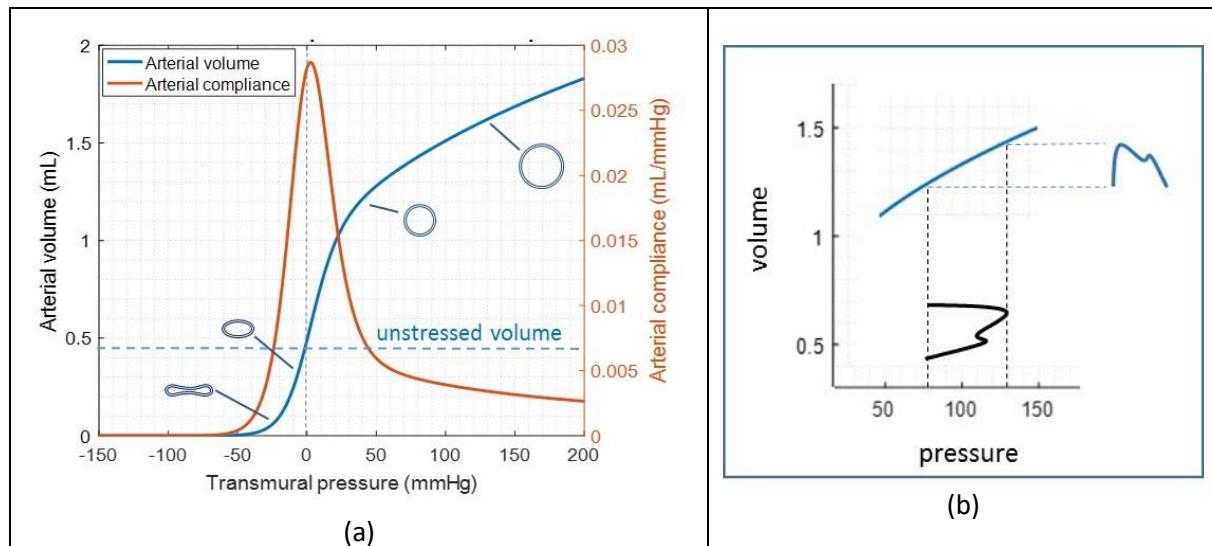


Figure 7-2 (a) Arterial volume and compliance versus transmural pressure. The shape of the cross section is indicated in the figure. The blue dashed line is the unstressed volume. (b) Subsection of (a). An arterial pressure waveform (black) and a corresponding volume variation of the artery (blue) are indicated.

7.2.2 Body locations for blood pressure measurement and procedures

For practical and physiological reasons, the clinician measures arterial blood pressure on a limited number of body locations. They are shown in Figure 7-3 (a).

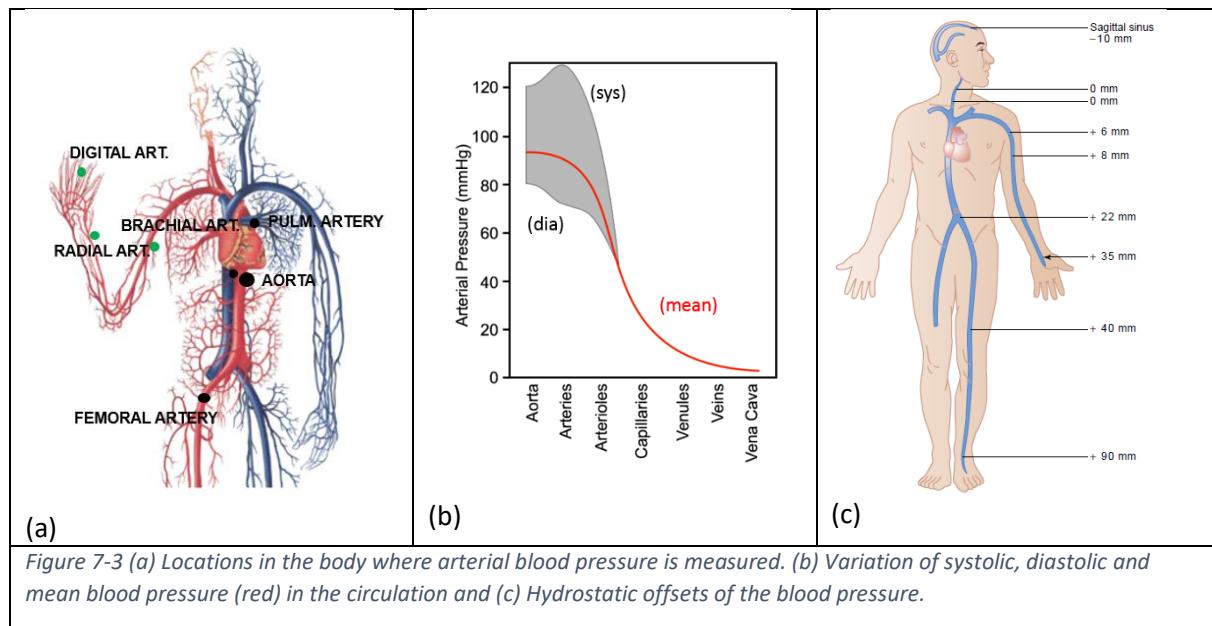


Figure 7-3 (a) Locations in the body where arterial blood pressure is measured. (b) Variation of systolic, diastolic and mean blood pressure (red) in the circulation and (c) Hydrostatic offsets of the blood pressure.

Several sites can only be reached by invasive techniques. The central aortic blood pressure (i.e. a measure of the afterload of the heart) is the most relevant parameter for clinicians. The central arterial blood pressure is measured in the aorta or aortic arch. Other locations for invasive blood pressure measurements are the subclavian artery, femoral artery and radial artery. These locations are used because of the easier access. The most frequently used site for invasive blood pressure is the radial artery, it is close to the skin, close to the periphery and has lowest risk for serious complications (bleeding, thrombosis). Pressures in the pulmonary circulation are measured after the right heart in the pulmonary artery. This requires special catheters that must pass the right heart.

before they reach the pulmonary artery. Locations for non-invasive blood pressure measurements are the brachial artery, digital artery (finger) and radial artery. The location at the brachial artery is used most frequently for non-invasive measurements. The location is practical, and the position is close to the height level of the heart. Therefore, errors due to hydrostatic pressure offsets can be minimized. It was shown in the physiology lecture that blood pressure varies in the arterial tree, especially systolic pressure and the waveform are affected (see Figure 7-3 (b)). Preferably the location of measurement should be close to the heart to minimize differences with the central blood pressure. Furthermore as was discussed during the treatment of the Bernoulli equation hydrostatic effects on blood pressure are large (see Figure 7-3 (c)) and cannot be ignored. Right atrial blood pressure is the pressure reference site for blood pressure. The gravitational effect is the cause that blood pressure measurements depend on posture. When standing every centimeter of height difference between the heart and body position corresponds with 0.75 mmHg measurement error. The measurement errors can easily be as large as tens of mmHg.

Note that blood pressure is not constant, varies with time, varies over the body, depends on posture, depends on breathing effort and physical activity and stress and depends on the time of the day (circadian rhythm). Blood pressure measurements should preferably be done after a period of 5 minutes of rest at a fixed position, posture and time or at multiple times all other factors being the same. The best body position is the supine position, laying on bed and measuring after a few minutes of rest. When sitting or standing the sensor location should be on the level of the heart or corrections due to height difference need to be applied. Venous blood pressure is measured during critical care in the ICU and invasive measurements are required. The most popular location is the subclavian vein, vena cava or the right heart location (Central Venous Pressure).

The most used methods for invasive- and non-invasive blood pressure measurement are described in the following sections.

7.3 Blood Pressure Measurement

The measurements that are most frequently used are shown in the diagram of Figure 7-4. As mentioned before blood pressure is a force per unit area. High-fidelity continuous blood pressure measurements measure the internal pressure in the artery. This requires invasive techniques; devices are placed by surgical techniques in the artery. This is the gold standard for blood pressure. Note that this procedure can be harmful for the patient, poses risks for complications and is typically done when there is a need for high-accuracy beat-to-beat blood pressure data. The benefit for the patient should outweigh the risk of the procedure. Invasive blood pressure is typically measured during emergency care, higher level surgery and care in the ICU. Recently miniaturized implantable devices have been developed, they are rarely used and pose risks for the patients (thrombosis). Implantable devices are not discussed in this course. For lower risk situations in the hospital or for use at the general practitioner or home, non-invasive methods are available. Since the true blood pressure cannot be sensed non-invasively, most of the non-invasive methods are *indirect* measurements. In most techniques a counter pressure is applied on the tissue around the artery and secondary effects are used to estimate blood pressure. The non-invasive methods are described after the invasive blood pressure measurements.

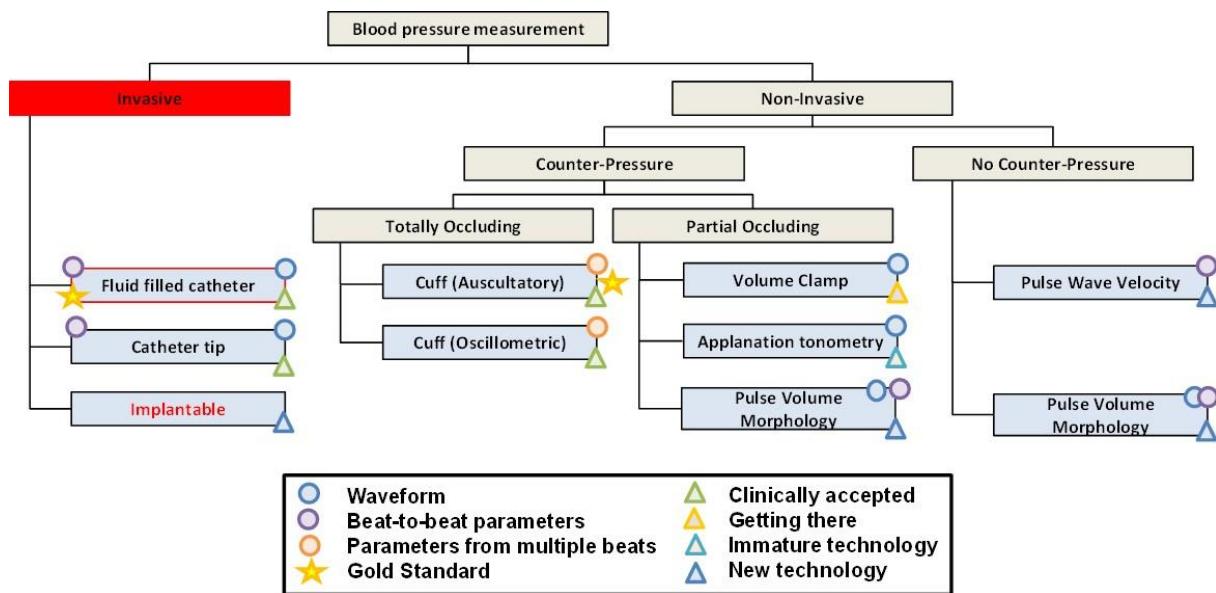
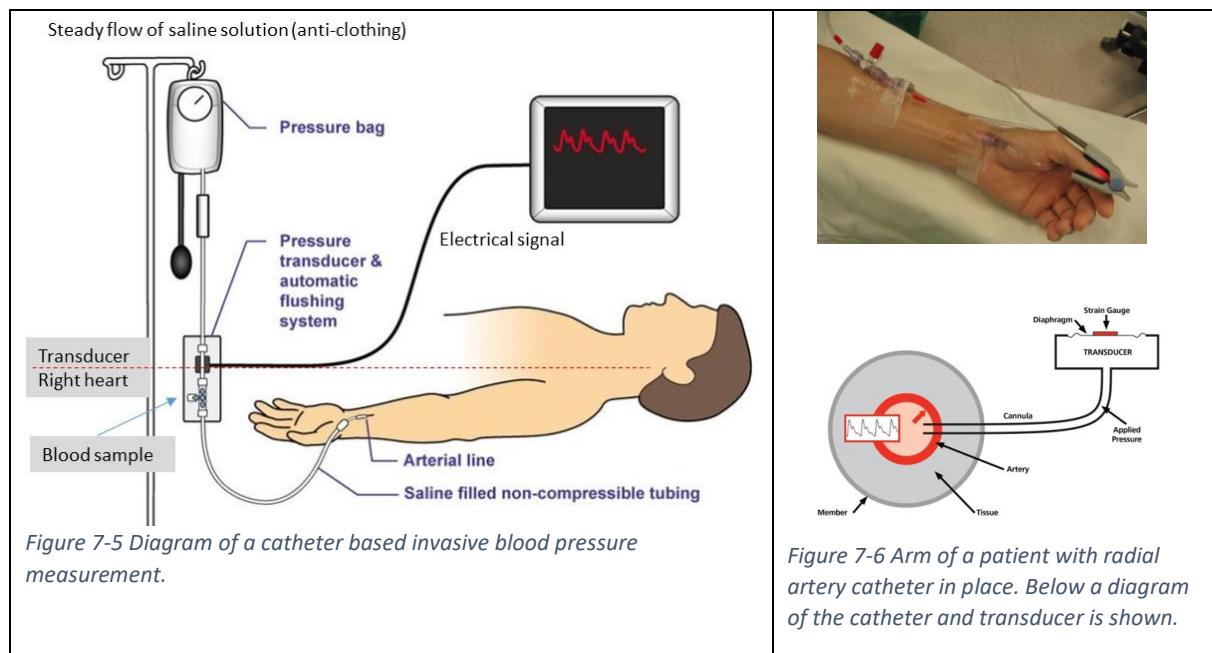


Figure 7-4 Diagram of the most frequently used blood pressure measurement techniques divided in invasive and non-invasive ones and techniques that use counter pressure or not. The type of measurement parameters, its clinical acceptance and gold standard techniques are indicated.

7.4 Invasive Blood Pressure

Invasive blood pressure is sensed in the lumen of a large artery. This requires a percutaneous procedure. The radial artery procedure is described below. A patient lies on a bed with the arm stretched in a horizontal position. A special inserter (needle) punctures through the skin and through the arterial wall and assists with the guidance and the insertion of a plastic catheter tube in the radial artery (see Figure 7-5, Figure 7-6). The catheter is then fixed with adhesive tape to the skin. The catheter is a flexible and somewhat stiff plastic tube with a diameter of a few millimeters with one or multiple lumens, each with a diameter in the order of a millimeter. A plastic bag filled with a saline solution is pressurized and subsequently connected to the catheter lumen and a small amount of fluid flows continuously into the artery. This flow is needed to prevent air bubble and blood clot formation in the catheter lumen. The direct access to the artery makes it possible to extract blood samples for from the artery for laboratory analysis of blood gases and chemical analysis of blood composition. Note that a catheter is a single-use device, once it has been in contact with blood it is not allowed to reuse the catheter (even after thorough cleaning). The arterial pressure pulse is transmitted via the fluid in the lumen to a pressure transducer (most often a strain gage transducer). The pressure sensor is not in the artery. Errors due to hydrostatic offsets can be large and the pressure sensor must be at the same height as the heart. The fluid filled catheter has compliance, inertance and resistive behavior. Hence it is a second order sensor system, pressure oscillations and/or under or overshoots at the sensor position occur frequently, the signal can be strongly distorted leading to errors in systolic and diastolic pressure measurement. There are procedures to optimize the damping parameter to obtain optimal measurement results. The preparation and execution of the invasive measurement requires special care. When these procedures are not done properly errors in the blood pressure measurement can be substantial (much larger than 10 mmHg).



Before starting the measurement, the zeroing and leveling procedure must be performed. The zeroing procedure exposes the sensor to atmospheric pressure and the reading is zeroed. Thereafter the leveling procedure is started. The position of the pressure transducer should coincide with the phlebostatic axis which corresponds roughly with the position of the right atrium. This is the accepted reference position for blood pressure measurements. An error of a few mmHg is acceptable, this corresponds to an error of a few centimeter. This height error is easily made, it is difficult to estimate the correct transducer position. Sometimes the height correction is omitted, sensor position errors of 10 cm or more are not uncommon. It is important to realize that it requires great care and skills to prevent errors in invasive blood pressure measurement. Studies have shown that in a large fraction of the measurements (~30%) settings are not optimal.

A pressure pulse travels through the catheter to the transducer position (see Figure 7-7 (a)). The catheter is an elastic tube with a fluid filled lumen. The catheter has a compliance C_c , a flow resistance R_c and fluid inertance L_c . The sensor is of the strain gage type and has a compliance that is larger than that of the catheter. The fluid filled catheter can be modeled as a short length tube and the schematic of Figure 7-7 (b) can be used to model the transfer characteristics of the tube. A normalized transfer function for a series LRC circuit is shown in Figure 7-7 (c). As was described in the lecture on fluid dynamics, oscillatory behavior can be observed when the damping parameter ζ is smaller than 1. The damping parameter ζ is equal to:

$$\zeta = \frac{R_c}{2} \cdot \sqrt{\frac{C_c}{L_c}} \quad 7.4-1$$

A second important parameter is the resonant frequency, it is equal to:

$$f_r = \frac{1}{2\pi\sqrt{L_c C_c}} \cdot \sqrt{(1 - \zeta^2)} \quad 7.4-2$$

The damping parameter depends on the stiffness of the tube, the compliance of the sensor and on the lumen diameter (influence on resistance, inertance). The magnitude of this damping parameter depends on the catheter design parameters. There is a tradeoff between bandwidth and signal distortion by oscillatory under and overshoot of the pressure signal. An optimal value for ζ is 0.7. Unfortunately, the parameter ζ is often much smaller or larger than this value and significant ringing and distortion is observed in the measured pressure signal. A typical resonance frequency is in the order of 20 to 50 Hz. The bandwidth of the signal is limited, it is of the order of 15Hz.

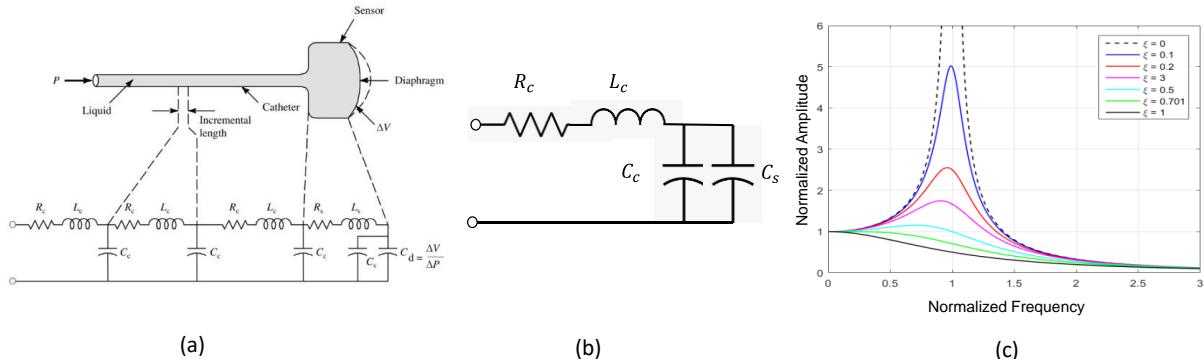


Figure 7-7 (a) Physical model of a catheter-sensor system (Webster). (b) Schematic model of an ideal catheter. (c) Frequency response of a catheter system with variable damping parameter.

The interaction between resonance frequency and damping parameter is illustrated in the diagram shown in Figure 7-8 (a). The region with the best signal fidelity is indicated (optimal response). The dashed square is the region where most of the clinical measurements are performed. It appears that in many cases the damping and resonance parameters are not optimal, as a result the pressure waveform is distorted, systolic and diastolic pressures deviate from the true arterial values. The impact can be severe as is illustrated in Figure 7-8 (b). When the system is overdamped (too large ζ) the pulse pressure amplitude and systolic pressure are underestimated. Overdamping can occur when there is a blood clot or air bubble in the catheter or when the catheter has a kink. In the case of underdamping the systolic and pulse pressure are overestimated and diastolic pressure underestimated. The damping parameter of the system must be optimized; large errors can occur when this is not done. When done properly the invasive measurement is the best measurement available. Unfortunately, as mentioned before measurement errors in invasive blood pressure are common.

During clinical use a simple method is used to estimate the damping behavior of the system. It is the fast-flush test. A square pressure wave is applied to the catheter-transducer system by a switch placed near the transducer (see Figure 7-9 (a)). The resulting pressure signal is monitored and analyzed (see Figure 7-9 (b)). When two to three oscillations are observed after the pulse, damping is optimal. For the overdamped case there is one or no oscillation. The overdamping can be caused by an air bubble or blood clot in the catheter, or when the catheter is kinked. Flushing with a saline solution removes the blockage of the catheter. Underdamping occurs when there are multiple oscillations observed. The origin could be related to the intrinsic properties of the catheter, too large length of the catheter or too many stopcocks in the catheter. Underdamping can be corrected either by electronic means (filtering) or by increasing the lumen resistance by for instance applying a mechanical clamp to locally reduce the lumen diameter.

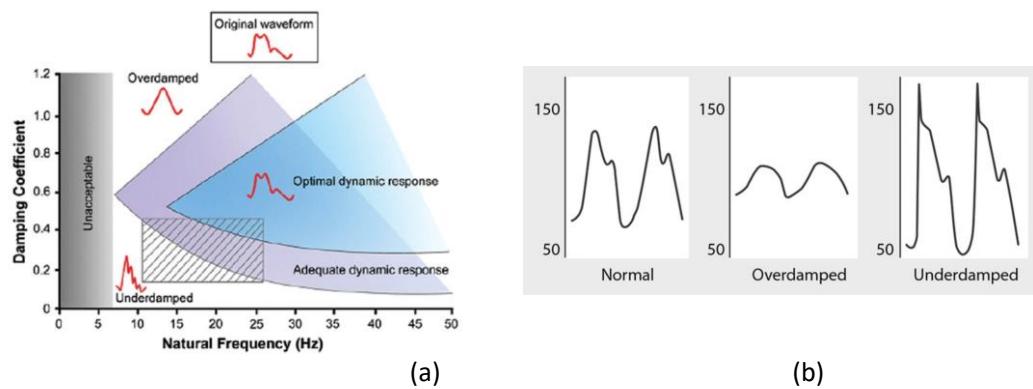


Figure 7-8 (a) Diagram showing the optimal region of damping parameter and resonance frequency. The dashed box is the area that corresponds with typical clinical use. (b) Pressure waveforms for normal, overdamped and underdamped catheters.

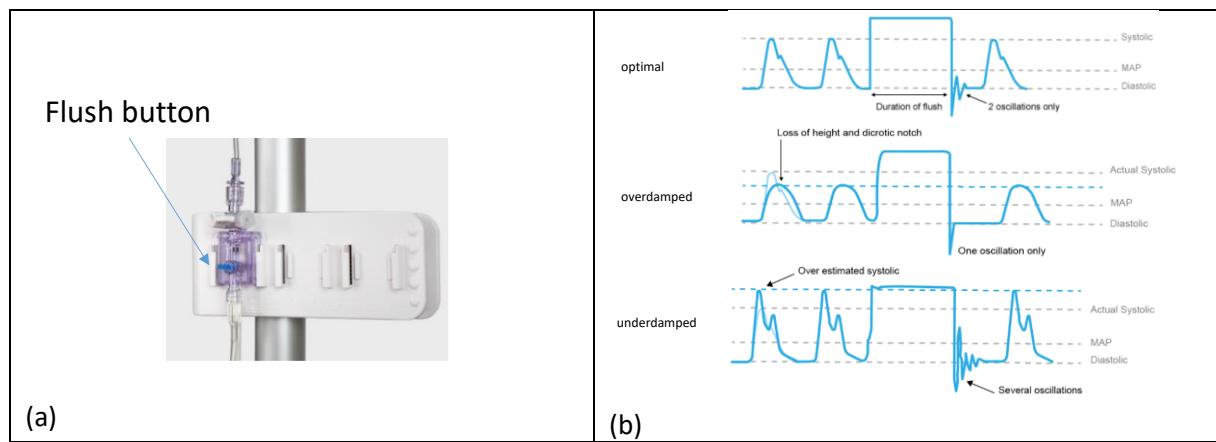


Figure 7-9 Fast flush button on arterial line transducer and response of the system on a square wave pressure pulse for normal, overdamped and underdamped conditions.

Further error sources are movement of catheters due to high velocity blood pulses and location and orientation of the catheter tip in the artery (too close to the arterial wall, direction of lumen opening with respect to the flow direction). When the zeroing, leveling and fast flush test have been performed blood pressure measurements are accurate to ± 5 mmHg. The fast-flush test may have to be repeated many times as clot formation and bubble formation may occur again after some time.

Measured radial blood pressure pulses are shown in Figure 7-10 for a 25 year, 50 year and 75 year male adult. This pressure pulses shows the forward and reflected pressure pulses and increase in pulse pressure and blood pressure with age. The bandwidth is limited to about 20Hz. Note that the waveforms deviate a lot from the predictions of the 2-element Windkessel model.

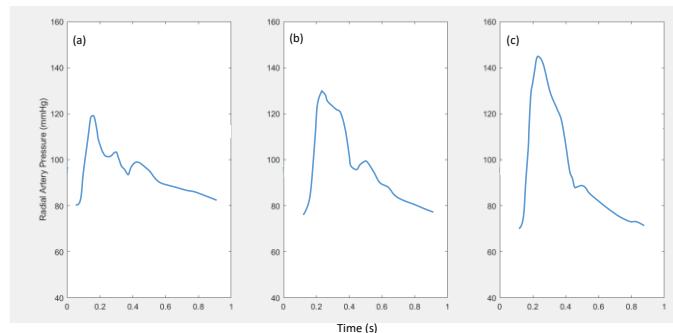


Figure 7-10 Recorded radial artery pressure waveforms of (a) 25 year (b) 50 year and (c) 75 year male adults.

Catheters with a MEMS pressure transducer placed at the tip offer higher fidelity measurements and are discussed in the following section.

7.4.1 Catheters with MEMS transducers

The conventional catheter has limitations in bandwidth and signal distortion. Furthermore, corrections for hydrostatic errors must be made and this procedure is prone to errors. For some use cases measurements with a much higher bandwidth may be needed. In the Cathlab a smaller size sensor is needed when pressure measurement(s) in small coronary arteries is needed. These sensors are used during cardiac characterization by a cardiologist or measurements of the cardiac performance in patients with heart diseases. In these use cases the limitations caused by the fluid filled catheter lumen cannot be overcome. A solution is to place a small transducer in the catheter tip. Due to the continuous miniaturization of microelectronics and the development of MEMS technology ultra-small high-fidelity pressure transducers have become available. An example is shown in Figure 7-11. In Figure 7-11 (a) a diagram of a catheter with a MEMS pressure transducer in the tip of the catheter is shown. The transducer is separated from blood by a thin membrane, the other side of the pressure sensor membrane is at atmospheric pressure. The sensor measures the gauge pressure. A photo of a tip of a Millar (leading supplier) catheter and the MEMS transducer are shown in Figure 7-11 (b). The dimensions are in the order of a millimeter. There are even smaller diameter catheters as is shown in Figure 7-11 (c) where a catheter developed for use during coronary events in the Cathlab is shown. This small and very flexible catheter can be inserted into coronary arteries via the femoral artery and can reach smaller blood vessels in the coronary system. The bandwidth of a MEMS catheter tip transducer can be as high as 1000 Hz. Errors due to hydrostatic effects due to different position of the sensor pressure measurement site are absent. The most important sources of error are movement of the catheter during pulsatile flow, orientation of the sensing element and proximity of the arterial wall. The cost of the catheter is around 500 to 1000 Euro and are much too high for routine use (reuse of catheters is forbidden).

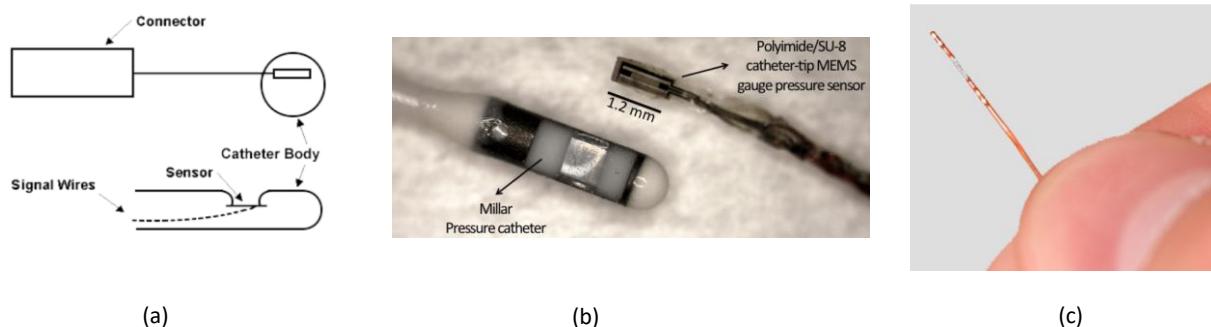


Figure 7-11 Catheter tip MEMS pressure transducer for measurement of gauge pressure. (a) Diagram of the sensor (b) photo of the sensor element and (c) photo of a micro catheter designed for coronary artery blood pressure measurement.

The aortic blood pressure of a 30 kg swine measured with a Millar MEMS catheter is shown in Figure 7-12 (a) and is compared with its femoral blood pressure measured with a conventional fluid filled catheter (Figure 7-12 (b)). There is a small offset in pressure due to a height difference between the sensors. The pressure waveform in the aorta shows many fine details and clearly shows the aortic notch. The measured pressure waveform in the femoral artery is severely distorted due to the limited bandwidth and second-order sensor characteristics plus the arterial wave distortion due to wave reflections. The limited bandwidth of the catheter and transducer acts as a low pass filter and impact on the wave form is large.

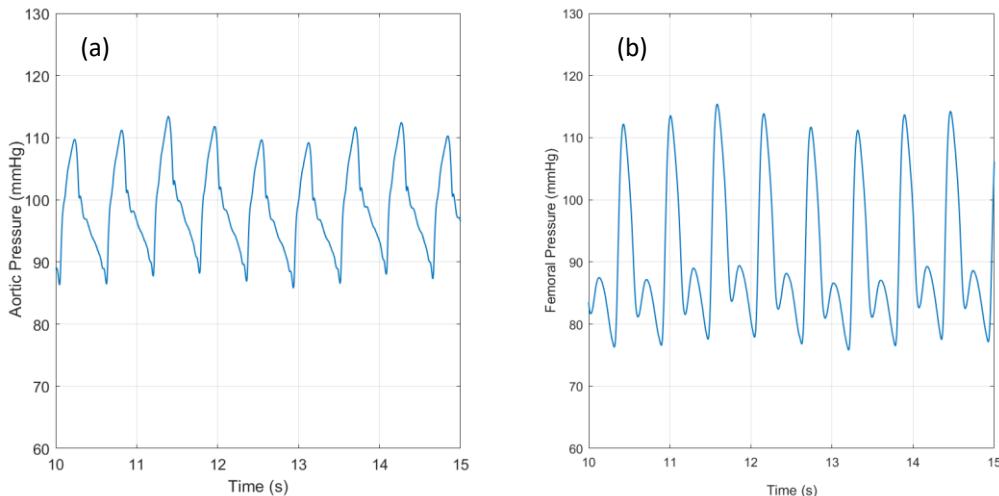


Figure 7-12 Plots of the aortic blood pressure (left) and femoral artery pressure (right) of a 30 kg swine model versus time. The aortic pressure was measured with a Millar catheter (bandwidth 1 KHz). The femoral pressure was measured with a conventional fluid filled catheter.

7.4.2 Summary Invasive Blood Pressure

Invasive blood pressure requires percutaneous placement of catheters in the lumen of a large artery. This is the gold standard for blood pressure measurement. Invasive measurements pose a risk for the patient and can in some cases cause severe complications for the patient. Therefore, the clinician must balance the risk of the procedure versus the benefit of a continuous high-fidelity blood pressure measurement. Typically, invasive measurements are done when the risk for the patient is high and continuous high-fidelity measurements are needed, for example during critical care in the ICU or some types of surgery. Critical care patients in a high-risk environment of the ICU have one or more invasive blood pressure measurements (arterial and venous blood pressure). A benefit is that it is possible to take blood samples which is often required for laboratory measurements. During surgery invasive blood pressure is measured in around 10% of the patients, mostly during long duration surgery, cardiac surgery and surgery where large blood loss can occur. Non-invasive blood pressure measurements are used for the majority of the patients. The most used techniques are discussed in the next chapter.

7.5 Non-Invasive Blood Pressure

Blood pressure is defined as a force per unit area, this can only be measured with invasive methods. Non-invasive measurements of blood pressure use indirect methods and are less accurate and less reliable than the invasive measurements. However, often the precision is adequate, and trends and changes can be measured reliably. Note that changes in vital sign parameters are often used by clinicians to guide therapy. A non-invasive measurement is preferred for routine care or when the risk of an invasive measurement is not warranted.

The most used methods apply an external counter pressure to the tissue surrounding the artery where the blood pressure is measured. In most cases the counter pressure is applied by an inflatable cuff that is wrapped around a body extremity. The most used location is the upper arm, the measurement at this location is discussed below. The pressure in the brachial artery is estimated using a measurement of a parameter that is related to the arterial blood pressure. There are several methods that will be described later. In Figure 7-2 a volume-pressure plot of the brachial artery was shown. This is shown in more detail in Figure 7-13. In the normal situation the tissue pressure is atmospheric, and all pressure falls over the arterial wall. The blood pressure is pulsating between

diastolic and systolic blood pressure. This causes a pulsating arterial volume, moving arterial wall and finally this motion is transferred via the arm tissue to the surface of the arm. A small pulsating change of the arm volume can be measured (see Figure 7-13 (a)).

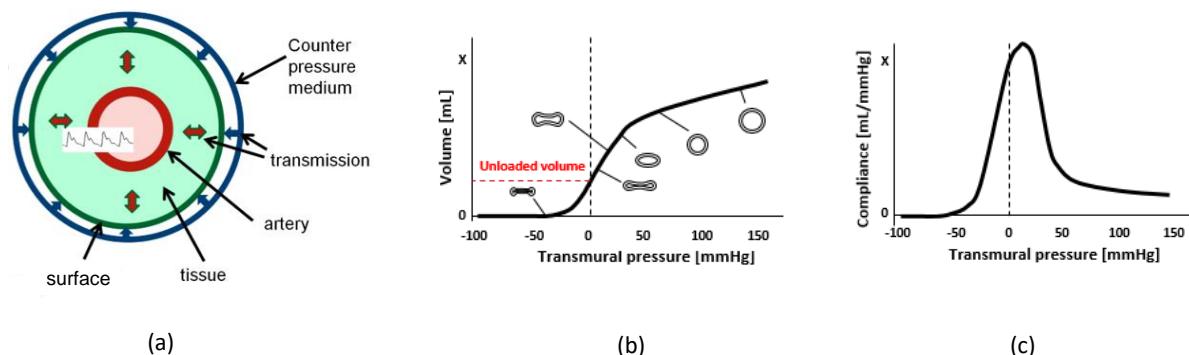


Figure 7-13 (a) Diagram of an inflatable cuff wrapped around a body extremity and indication of volume pulsations of the artery and tissues. (b) Diagram of the arterial volume versus transmural pressure and indication of the arterial cross section. (c) The arterial compliance versus transmural pressure.

When the cuff is inflated the counter pressure increases, the transmural pressure over the arterial wall decreases and the arterial volume decreases (see Figure 7-13 (b)). It should be noted that although the volume of the artery decreases the arterial volume pulsations increase until a maximum is reached around zero transmural pressure. Both the arterial compliance and the volume pulsation reach a maximum value at this cuff pressure (see Figure 7-13 (c)). A further increase of the external pressure reduces the arterial diameter until the artery is fully collapsed. Hence the volume pulsations become smaller and finally blood flow through the artery is blocked at a cuff counter pressure that is much larger than systolic blood pressure. The arterial volume changes induced by cuff pressure enable estimates of blood pressure. Different criteria and signals are exploited. The first attempt to estimate systolic blood pressure using cuff inflation was done by the Italian physician Riva-Rocci at the end of the 19th century.

7.5.1 Riva-Rocci Method

In the 19th century the inflatable cuff was developed for occluding blood vessels during surgery. The higher the blood pressure, the higher the cuff pressure required to occlude the artery. Riva-Rocci, an Italian physician, combined this inflatable cuff and a mercury manometer and developed a method to estimate blood pressure (see Figure 7-14). The pressure in the cuff where the radial artery pulsations disappear was assumed to be equal to the systolic blood pressure. The criterion to relate blood pressure with external cuff pressure was the disappearance of radial artery pulsations. Note that in reality this cuff pressure is larger than the real systolic pressure! Pulsations are finite when the cuff pressure is equal to the systolic pressure. Pulsations can be measured either manually or with a device. Although this was a first step the estimate of blood pressure, its value is not accurate as arterial pulsations disappear at higher cuff pressures than systolic pressure. Moreover, only the systolic pressure was estimated. The precision is not good as manual palpations are also notorious unreliable. Hence there was a need to come up with a method that was more sensitive, accurate and that could measure both diastolic and systolic blood pressures.



Figure 7-14 Riva-Rocci and a schematic drawing of a cuff based blood pressure measurement based on his criterion and a photo of an early mercury sphygmomanometer.

7.5.2 Korotkoff Auscultatory Method

Korotkoff discovered that sounds are emitted from the artery and arm distal from the cuff when the cuff pressure is varied from high (i.e. greater than systolic pressure) to low pressures (i.e. below diastolic pressure). These sounds are generated by blood flow in the artery under the cuff due to heart beats. He placed a stethoscope distal from the cuff at the inner side of the elbow and observed a reproducible audible sound pattern (see Figure 7-15 (b)) when the cuff pressure was varied from above systolic pressure to below diastolic pressure. These sounds are called Korotkoff sounds after the discoverer. Korotkoff was a Russian surgeon who lived from 1874 to 1920.

When a cuff is inflated above systolic pressure (~30 mmHg above systolic pressure) and the cuff is slowly deflated, audible sounds are observed. Every heartbeat generates a sound in the form of beats. A first sound is detected by the human observer at a fixed value of the cuff pressure (sound I), sounds become louder (sounds II), they reach a maximum loudness (sound III) and finally sound levels decrease (sound IV) and become very weak (sound V) fade away at a fixed value of the cuff pressure.

The exact origins of the Korotkoff sounds are still under debate. The sounds are related to (distal) blood flow, turbulence, flow and shape instabilities in the collapsible brachial artery and possibly by arterial wall movements from the collapsed state to open state. The processes emit sounds in the audible range. The interpretation of the experimental results is shown in Figure 7-15 (c) where arterial and cuff pressure are plotted together with the occurrence of the Korotkoff sounds. At high cuff pressure, the artery is fully collapsed and there is no blood flow, no sounds can be observed. When the cuff pressure is equal to systolic pressure, the artery opens slightly, and blood starts to flow through the artery for the first time. Korotkoff sound I is believed to correspond to the first opening of the artery. The corresponding cuff pressure is identified as systolic arterial pressure. When the cuff pressure is further reduced, the arterial volume impulses increase, blood flow increases and sound intensity increases (see Figure 7-15 (c)). With a further reduction of the cuff pressure the period of collapse becomes shorter, the lumen diameter increases, and sounds become weaker. Finally, when the cuff pressure is close to the diastolic pressure the artery does not collapse anymore, turbulence and wall motion are almost gone and audible sounds are very weak, below the audible threshold of the observer. The cuff pressure where sounds cannot be observed by an observer is assumed to correspond with the diastolic pressure. This interpretation is simple, the interpretation resonates with clinicians and this method is accepted almost universally during clinical education. The physical reality is different from the perception of the clinicians, with a sensitive

microphone sounds can be detected above and below systolic and diastolic pressures. In short, the physical phenomena are very complex and there is still debate on the origin of Korotkoff sounds and accuracy of the method. The method is empirical, there is no theory that supports the criterion.

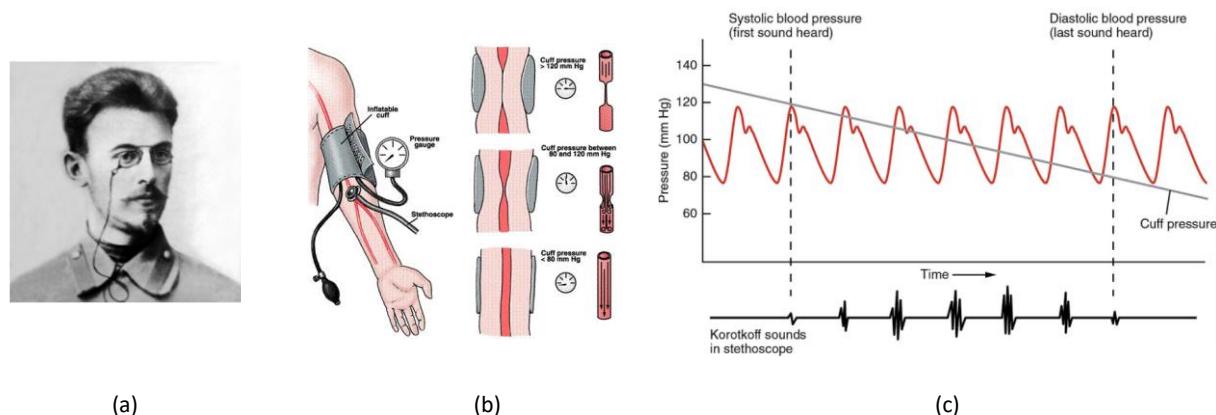


Figure 7-15 (a) Korotkoff the inventor of auscultatory blood pressure. (b) Diagram of the auscultatory measurement and occlusion of the artery during cuff inflation. (c) Diagram that shows the relation between arterial blood pressure, cuff pressure and detection of Korotkoff sounds.

This method is called auscultatory non-invasive blood pressure. The criterion for systolic pressure is the cuff pressure when the first sound is detected by the observer. The criterion for the diastolic pressure is the cuff pressure when the sounds disappear. A typical deflation rate is between 5 mmHg and 10 mmHg per second.

The auscultatory measurement is considered by cardiologists as the gold standard of non-invasive blood pressure measurement. This is remarkable considering the errors that have been reported and the empirical nature of the measurement. Some clinicians have a higher trust in this measurement than in the invasive one. The reason might be that invasive measurements can have large errors when the procedures are not executed properly, and beat to beat blood pressure is often highly variable (this is a real effect but is ignored in clinical practice when non-invasive measurements are done). The Korotkoff method requires great skills of the operator. The method also has some systematic flaws. There are often relatively large deviations between results from the non-invasive method and invasive ones especially for certain patient groups (mean pressure error > 10 mmHg). The physiological background of the method is dubious. There are concerns about the methodology. First the method is operator dependent. Accuracy and precision depend on the hearing ability of the operator, depends on background noise, resolution depends on the deflation speed. Furthermore, measurements with a sensitive microphone have shown that the sounds do not appear or disappear suddenly, weak sounds are measured with a microphone above systolic and below diastolic pressures. Finally, there is sometimes an auscultatory gap, sounds are missing at cuff pressures between diastolic and systolic pressure. Measurement have large errors for certain patient groups (for instance during pregnancy). Pileup of blood in the arteries and veins distal from the cuff reduce blood flow in the arm during the measurement and this may influence the measurement results. Furthermore, it is difficult to automate this measurement, and there is a demand for non-invasive automatic measurements of the blood pressure. The oscillometric non-invasive method is easy to automate and is discussed in the next section.

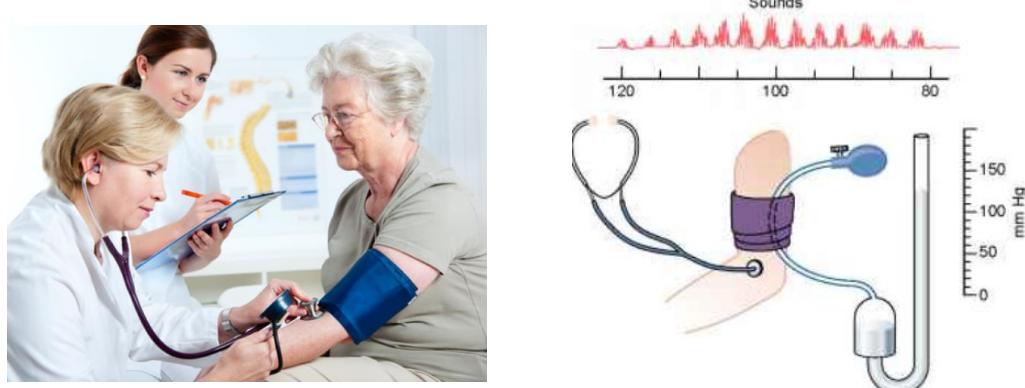


Figure 7-16 Methodology used during auscultatory blood pressure measurement. During cuff deflation a clinician notes the cuff pressure at which sounds appear and disappear.

7.5.3 Oscillometric Non-Invasive Blood Pressure

The oscillometric blood pressure measurement technique is based on an early finding in the 19th century that the mercury column starts to oscillate up and down for cuff pressures between diastolic and systolic pressures. As described before changes in arterial volume due to the pulsatile arterial volume are transmitted via the arm tissue to the arm surface and finally to the cuff. The air volume of the cuff bladder oscillates, and the resulting small cuff pressure changes can be measured (Boyle's Law). This is shown in Figure 7-17.

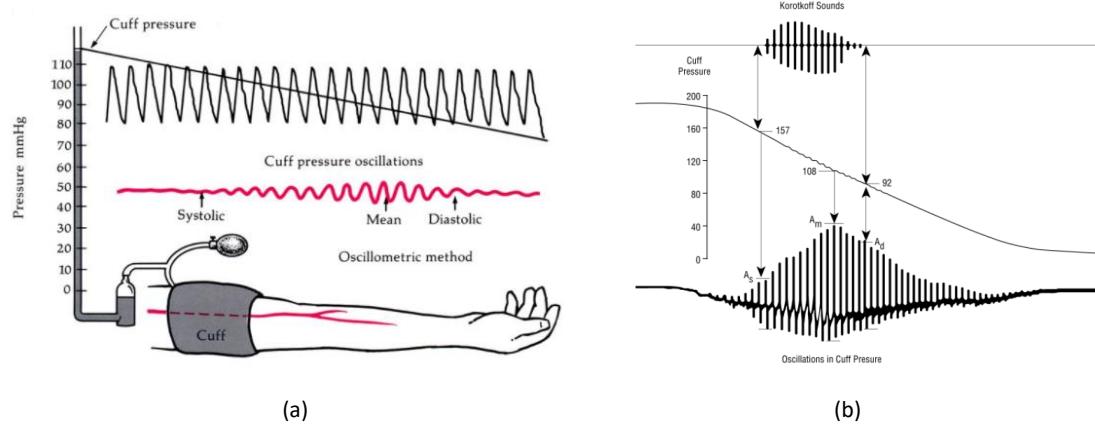


Figure 7-17 Diagram of an oscillometric blood pressure measurement (a) and Korotkoff sounds and cuff pressure and low-pass filtered cuff pressure (b). Note that the Korotkoff sounds are estimated and are for illustration purposes only.

During inflation or deflation, the cuff exerts a pressure on the arm surface and this pressure is transmitted via the arm tissue to the outer arterial wall. The pulsatile blood pressure in the artery varies between diastolic and systolic pressure. The transmural pressure of the artery changes due to changes in the cuff pressure, the resulting arterial volume changes depend strongly on the transmural pressure via the volume-pressure relation shown in Figure 7-2 and Figure 7-13. This is shown in some more detail in Figure 7-18. For a given transmural pressure pulse the corresponding change in arm volume is shown. For small cuff pressure the transmural pressure is large and the change in arm volume is small, hence cuff pressure oscillations are small. When cuff pressure increases the transmural pressure decreases and the change in arterial volume caused by blood pressure variations increases. The cuff bladder volume pulsations and related cuff pressure oscillations increase strongly. When the transmural pressure pulse includes the zero pressure point the volume oscillations and cuff pressure oscillations are largest. With a further increase of cuff pressure transmural pressures are negative and arm volume oscillations and related cuff pressure

changes decrease strongly. A typical cuff bladder volume is in the order of 200 ml, the arm volume changes are in the order of one milliliter. Hence the maximum cuff pressure oscillations are smaller than a few mmHg. The small pressure oscillations are superimposed on the larger slowly varying cuff pressure (see Figure 7-17 (b)). The cuff pressure ramp is a very low frequency signal and can be removed from the signal by a high pass filter (cut off frequency ~ 0.1 Hz). Often a band pass filter is used, this filter also removes high frequency electronic and pump noise.

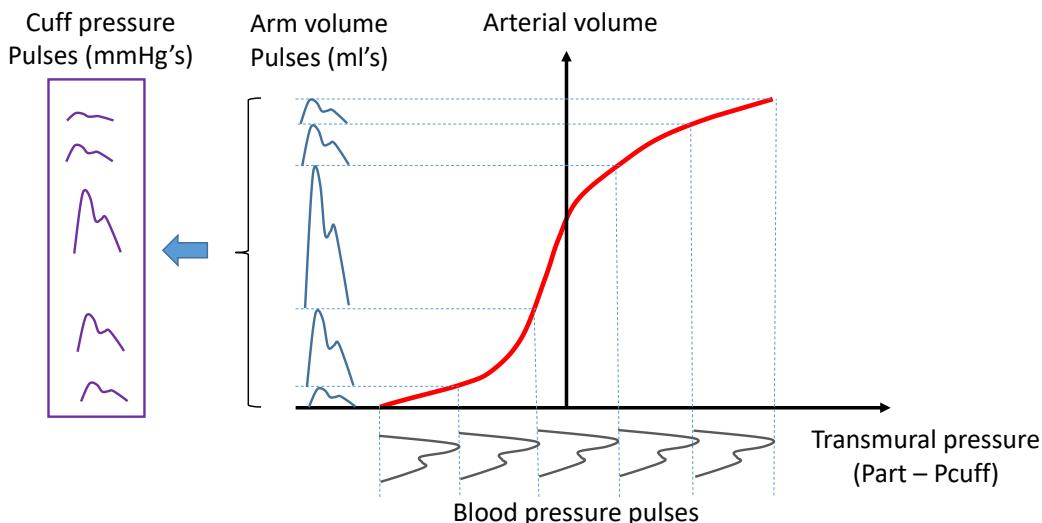


Figure 7-18 Diagram of the arterial volume-pressure curve including pulses of the transmural pressure, pulses in arm volume for a specific pressure pulse and finally cuff pressure oscillations.

A schematic diagram of idealized Korotkoff sounds and high pass filtered cuff pressure signals is shown in Figure 7-17 (b). The systolic and diastolic pressures determined from the auscultatory measurement are indicated in the figure. The cuff pressure oscillations extend beyond these pressures.

The criteria to extract systolic, mean and diastolic blood pressure from the cuff pressure oscillations were studied first by the group of Prof Geddes at Purdue University around 1980. In animal and human studies, they observed that in many cases, the systolic and diastolic pressure could be estimated with good accuracy from characteristic points of the normalized cuff pressure envelope. The diastolic and systolic pressure correspond with the cuff pressure at normalized pressure oscillation amplitudes A_d and A_s (i.e. normalized to the maximum cuff pressure amplitude). They found that most often the diastolic arterial pressure corresponded to a value of 80% of the peak value at the low pressure side of the envelope and that systolic pressure corresponded to the cuff pressure with a pressure oscillation amplitude of 55% of the maximum value at the high pressure side. They also found that the cuff pressure with the maximum pressure oscillation corresponds to a good degree to the mean arterial pressure. Note that this method is also empirical, works well in the normal blood pressure range on healthy volunteers and is based on correlations with clinical data. Please note that these characteristic normalized amplitudes are population averages, determined on populations of relatively healthy persons and young laboratory animals. In practice the characteristic values can be person specific and can even be time dependent.

The procedure to extract the arterial blood pressure values from an oscillometric method is shown in more detail in Figure 7-19 for a measurement during the cuff inflation. After the cuff is wrapped around the upper arm the cuff is inflated and cuff pressure (blue line left figure) is recorded with a pressure sensor with high accuracy ($< 1\%$ FS) and resolution (< 0.1 mmHg). Data is sampled at a

frequency around 100Hz. Pressure oscillations are extracted after high pass filtering of the acquired pressure time series (blue pressure oscillations, right figure). The resulting pressure oscillations are converted to a pressure envelope (i.e. magnitude of the pressure oscillation as function of cuff pressure) where the horizontal axis is the low pass filtered cuff pressure and the y-axis corresponds with the amplitude of the pressure oscillations. The envelope is constructed by the magnitude of the pressure oscillation, i.e. the difference between a maximum and the adjacent minima. From the measured envelope the diastolic, systolic and mean pressures are determined by using the characteristic ratio method.

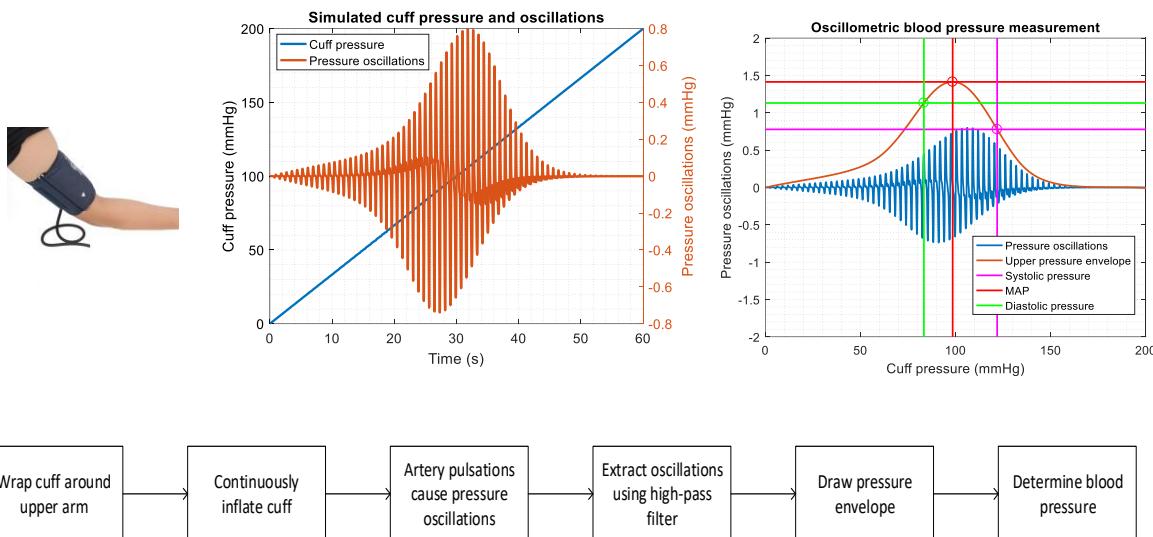


Figure 7-19 Procedure to extract blood pressure values from an oscillometric blood pressure measurement.

Note that different manufacturers can use other measurement procedures to obtain better agreement with the gold standard reference measurement. Most often the pressure oscillations are measured during a step wise deflation. The advantage is that pressure amplitudes can be determined with better accuracy. However, the measurement time is longer and the method is more obtrusive to the patient. Furthermore, pile up of blood in the artery and veins distal of the cuff can cause systematic errors.

The physiologic basis for the oscillometric NIBP method is based on the Volume-Transmural Pressure relation of the artery. The inference of blood pressure from characteristic values is empirical and is also person dependent. Nevertheless, the accuracy of blood pressure values is comparable to that of the auscultatory measurement in test groups used for qualification of the product. Unfortunately these test groups are relatively small, not diverse and there too few participants with hypo- or hypertension. The advantage of the oscillatory method is that it can be automated easily, and that the skill of the operator is less important. Therefore, this method is used widely in clinical practice. However, before a specific measurement device can be used in the clinic or even at home its accuracy has to be verified. The accuracy and verification methods are discussed in the next section.

7.5.4 Accuracy and Verification

There are several clinical verification tests for non-invasive blood pressure devices. There are documents in standards from the British Hypertension Society, the Association for the Advancement of Medical Instrumentation and the European Society of Hypertension International Protocol. Oscillatory blood pressure monitors must meet several criteria. To meet these criteria, devices must pass the accepted standards of the ESH International Protocol 2002 (IP1) or ESH International

Protocol 2010 (IP2) or achieve a minimum B grade for both systolic and diastolic measurements for the revised (1993) BHS protocol. The BHS protocol is shown in Figure 7-20.

The BHS protocol: "The basis of device evaluation is the comparison of blood pressure measured by the device being tested with measurements made by trained observers using a mercury sphygmomanometer and stethoscope to auscultate the Korotkoff sounds." A group of 85 test persons with a wide and specified range of blood pressures is selected for the test. *The reference method is the auscultatory method based on Korotkoff sounds.* Two independent trained observers determine blood pressure with the reference method. *The use of this indirect reference method is justified by clinical practice, clinical decisions for hypertension are made using the results of indirect auscultatory measurements.* Large deviations of the auscultatory reference measurement with invasive direct methods in significant number of test persons as well as the variation with time of blood pressure are mentioned in the standardization documents but are discarded by the standardization committee.

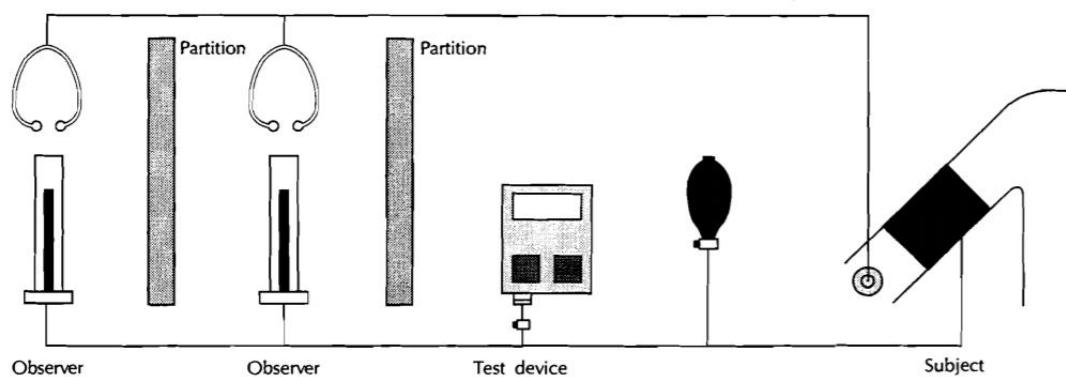


Figure 7-20 BHS protocol for NIBP device validation.

The results from the reference method and the device under test are compared and the data is divided in three groups with a difference between the means of 0 and ± 5 mmHg, 0 and ± 10 mmHg and 0 and ± 15 mmHg respectively. There are four classes (A, B, C, D) with A being best. The classification depends on the fraction of the measurements in the three groups. The AAMI recommendations are that the difference between the means shall be ≤ 5 mmHg and the standard deviation shall be ≤ 8 mmHg. This means that for a non-negligible fraction of measurements differences are larger than 20 mmHg. Furthermore, note that the number of test persons with hypo- or hypertension is relatively small.

A comparison between invasive radial artery pressures and oscillatory blood pressure values is shown in Figure 7-21. A group of more than 20000 patients was included in the study. For the normal range of blood pressure, the deviations are in line with the AAMI standard. For patients with severe hypotension or hypertension the deviations are systematic and unacceptably large.

There are many smaller studies that confirm these results. There are similar studies for the comparison of invasive and auscultatory blood pressure. Deviations larger than the AAMI values are observed in many studies. In summary the deviation between the oscillometric and auscultatory non-invasive blood pressure methods and direct invasive arterial blood pressure is often large, especially for patient groups with extreme values of low and high blood pressure. Changes in blood pressure and trends are more reliable than the absolute values and this is often more important in clinical practice.

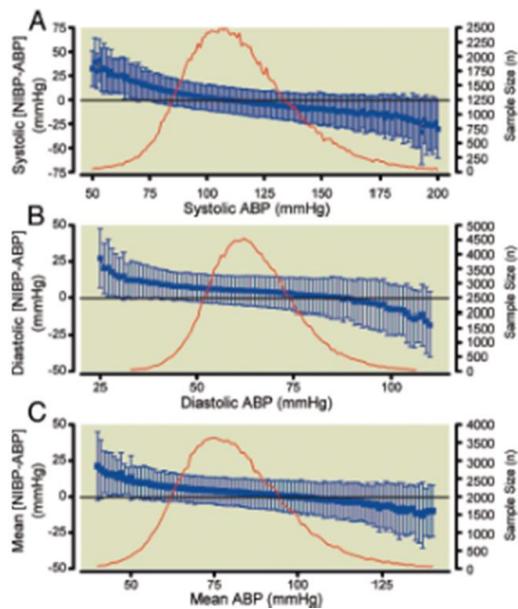


Figure 7-21 Measured differences of systolic, diastolic and mean blood pressures determined with oscillometric NIBP and direct radial artery measurements.(from Wax et al., Anesthesiology 2001, 115, pp973). The orange lines are the distributions of the number of patients and their blood pressure.

7.6 Continuous Non-Invasive Blood Pressure

The auscultatory and oscillometric measurements are intermittent measurements. The oscillatory method is integrated in patient monitors and can be performed at regular time intervals. A short interval is 5 minutes, the regular inflation of the cuff to high pressures is uncomfortable for the patient and can in some cases cause injuries. In certain conditions even a 5-minute interval may be too long. There is a need for less obtrusive and continuous non-invasive blood pressure measurements. There are three options which are described in the following sections.

7.6.1 Applanation Tonometry

Tonometry was developed initially to improve and quantify the manual palpitation of the arterial pulse. The applanation tonometer is the latest development and its aim is to record a continuous pressure waveform. In Figure 7-22 a diagram of the principle of operation of an applanation tonometer is shown. The elastic wall stress, transmural pressure and radius of the artery are related via the Laplace equation (see Figure 7-22 (a)). Normally the external tissue pressure is equal to atmospheric pressure. The arterial pressure is dropped completely over the wall. In other words, the arterial wall is loaded. The larger the radius, the smaller the transmural pressure and pressure drop over the wall (for a given wall stress).

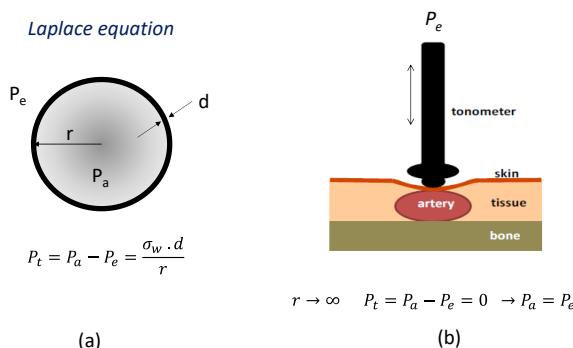


Figure 7-22 (a) Diagram of an artery and the Laplace equation. (b) Applanation of an artery.

When an artery is compressed and flattened (but not occluded) the local radius in the flattened part becomes very large and according to the Laplace law (Figure 7-22 (b)) the local transmural pressure is very small. The external and arterial pressure are almost equal in the flat part of the artery. The deflection of the sensing element or pressure signal from the sensor mounted on the actuator is proportional to the arterial pressure. The mean pressure is not measured well, and a pressure calibration is needed, this is done with a conventional oscillometric blood pressure device. A continuous waveform of the arterial pressure is obtained. This principle was studied in the 19th century by Marey (see Figure 7-23). He developed a sphygmograph (pulse writer) and was able to record the displacement of the skin at the radial artery and recorded a signal related to the arterial waveform.



Marey sphygmograph (1860)

Manual applanation tonometry

Tensys – Automated device

Figure 7-23 The Marey sphygmograph, manual and automated device.

For tonometry a large artery close to the skin is needed. Furthermore, the location of the artery must be fixed and should be supported by a hard tissue like a bone. The radial artery site is the best site, the carotid artery is also used. The radial bone is reasonably flat, and the radial artery is close to the skin and bone. Furthermore, the sensor should be small enough such that it covers only the flat part of the artery. The position of the sensor with respect to the artery is fixed during measurement and sensor and patient motion during measurement is not allowed. The procedure during manual tonometry (Figure 7-24) is illustrated with the aid of Figure 7-24.

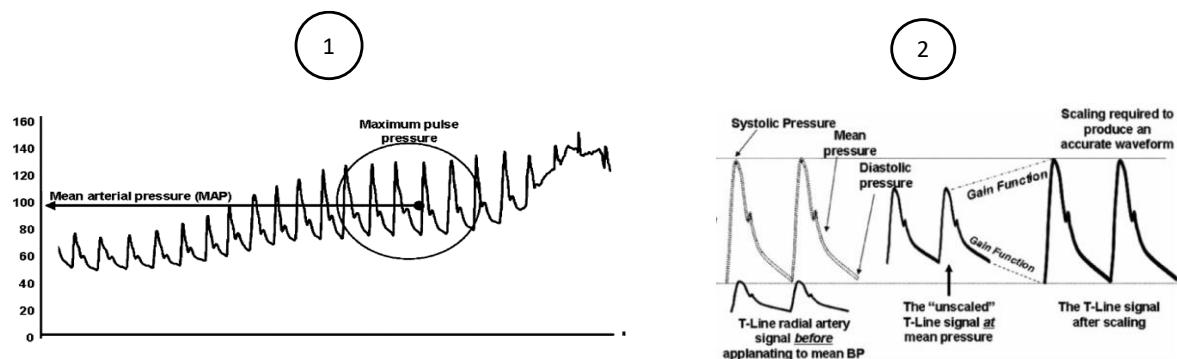


Figure 7-24 Procedures for manual tonometry.

The operator places the sensing element on the radial artery and adjust the holding pressure such that the maximum signal amplitude and correct waveform is recorded. This maximum signal amplitude corresponds with the optimal applanation. The measured signal is then scaled with the aid of a non-invasive blood pressure measurement. The procedure is difficult, depends on the skill of the observer and is not truly continuous as the signal is only measured when the operator is present. To circumvent the problems related to manual tonometry automated devices have been developed. An example is the Tensys device shown in Figure 7-23. The device has a sensor array and automatically selects the correct sensor position. The contact/holding pressure is regulated by a

servo control. Such systems are less affected by operator skills but still have several flaws. The radial artery is small and edge effects are not negligible. The viscoelastic properties of the skin and tissue distort the pressure waveform. Minute motion degrades the signal such that the data are not relevant anymore. Finally, during clinical validation studies, the results were not of sufficient quality to accept this method for large scale use.

7.6.2 Vascular Unloading Technique (Volume Clamp)

The vascular unloading technique was invented by Penaz in 1967 and refined in 1973. The principle of operation is illustrated using the diagrams in Figure 7-25.

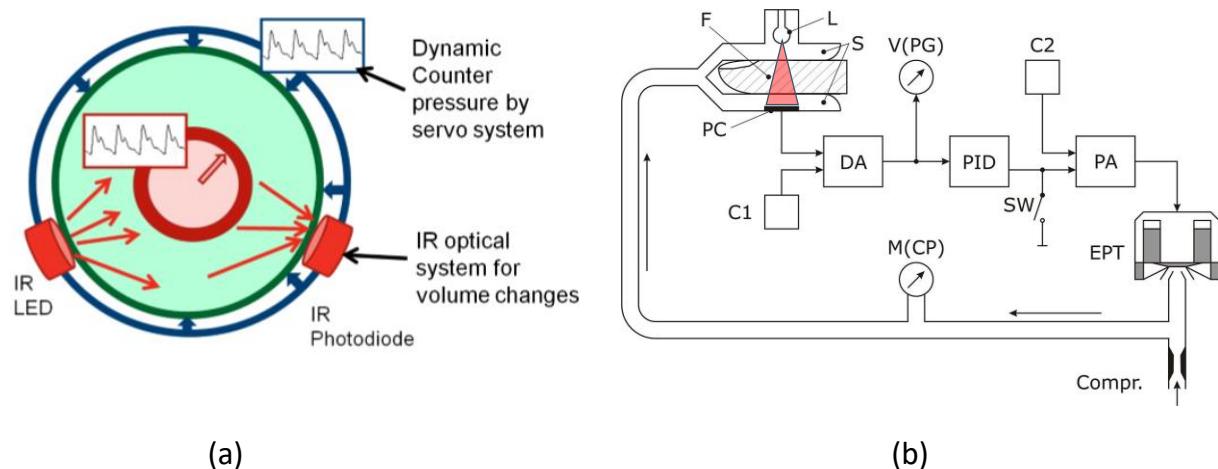


Figure 7-25 Diagrams to illustrate the vascular unloading technique.

A small cuff is placed around an index or middle finger. A servo-controlled system generates a dynamic counter pressure in the cuff to unload the artery wall, i.e. to maintain a zero transmural pressure (see Figure 7-27 (a)). The surface pressure is then equal to the arterial pressure and the volume of the artery is constant (volume clamp). The cuff pressure is then assumed to be equal to the arterial pressure. Note that there can be a DC offset between the arterial and cuff pressure because the method is capacitive and only pressure changes can be measured with some accuracy. The actuator includes a separate infrared photo-plethysmography (ppg) sensor that is sensitive to changes in the arterial volume (see Figure 7-27 (b)). The infrared LED emits continuous light that is detected by an IR photodiode. The PD signal consists of a DC part and an AC part, the AC part is related to the arterial volume changes. The control set point of the ppg signal (C_1) corresponds to the unloaded volume state of the artery. The PD signal and reference signal are subtracted, amplified (DA) and fed into a PID controller. Its output is the control signal of a power amplifier (PA) which drives a fast air pump. The pressure is controlled such that the PD signal is equal to the constant C_1 . In theory the cuff pressure should now be equal to the arterial pressure. However, a DC volume offset, and pressure offsets can be present. Regular (re)calibration is needed.

There are several manufacturers of volume clamp devices for continuous NIBP each with their own proprietary hardware and calibration algorithms (see Figure 7-26). A group from TNO in the Netherlands developed the Finapress device in the 1980's and a spin-off company of TNO (BMEYE) further developed the device. BMEYE was recently acquired by Edwards, the device is now marketed as Clear Sight. The calibration is done using the Physical™ method for initial and frequent recalibrations. A second manufacturer is CNS systems, its product CNAP uses other calibration methods based on a reference NIBP measurement. Finally part of the TNO group cooperates with Demcon to produce an alternative embodiment of the Finapress device.



Edwards Clear Sight



CNAP

Figure 7-26 Edwards Clear Sight and CNAP continuous NIBP devices based on the Penaz vascular unloading method.

The physics behind the measurement is relatively straightforward but similar to the oscillometric method there is no physical justification for the empirical calibration method to set the C1 constant and determine the DC offset pressure. The vascular unloading method has several flaws. The method to determine the unloaded volume is empirical and is not backed by physiology. It is difficult to find the correct set point for the unloaded volume and repetitive calibrations are needed during which the measurement is not available. A fast servo-controlled system using air to transmit pressure changes is very complex, it is on the boundary of being stable. Under and overshoots of the control process may be interpreted as real blood pressure values. The cuff pressure is continuously above venous pressure and blood will pile up in the finger distal to the cuff. This is uncomfortable to the patient but may also cause systematic measurement artifacts, the ppg signal is affected by the increased blood volume and flow reduction (flow reduces because the A-V pressure gradient becomes smaller, artery-venous reflex affects the vascular tone). There is also a concern during critical care when centralization of blood flow occurs or when certain anesthesia agents and vasopressors are used. Under these conditions the vascular smooth muscle tone changes rapidly, peripheral blood flow is reduced and ppg signals change dramatically and the control process may not be reliable anymore. Furthermore, the technique is not useful for patients with cardiac arrhythmias. Finally, the technique is uncomfortable for patients and may be used only for a limited time. Note that blood pressure is measured in the periphery and clinicians are interested in the central blood pressure or brachial artery blood pressure. This can be resolved by calibration and subsequent scaling using data obtained from an oscillometric NIBP measurement. Another method is to use a transfer function with parameters obtained from population averages (see Figure 7-27).

Volume clamp devices are commercially available for some time but are not used on a large scale yet. The costs per use case are high and the benefit is still unclear. However, the device has many attractive features and it is less obtrusive for the patient than the invasive measurement. When the high cost per use is reduced and the issues with vascular tone are resolved these systems may become attractive alternatives for invasive radial blood pressure measurements.

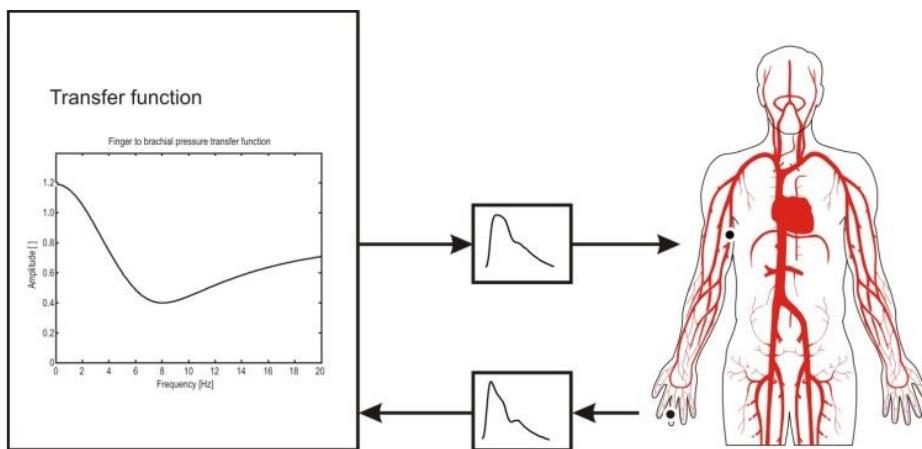


Figure 7-27 Diagram showing the transfer function method to obtain the brachial artery pressure from the finger pressure.

7.7 Continuous Non-Invasive Blood Pressure – Surrogates

As mentioned in previous sections there is an urgent need for a continuous non-invasive blood pressure measurement which is less obtrusive than the techniques described in the previous sections. There are several methods, the common factor is that parameters are measured that are even a worse surrogate for the true blood pressure than used in the previous measurements. The most popular method is based on the pressure dependence pulse wave velocity (PWV), the propagation velocity of pressure pulses in the arterial system.

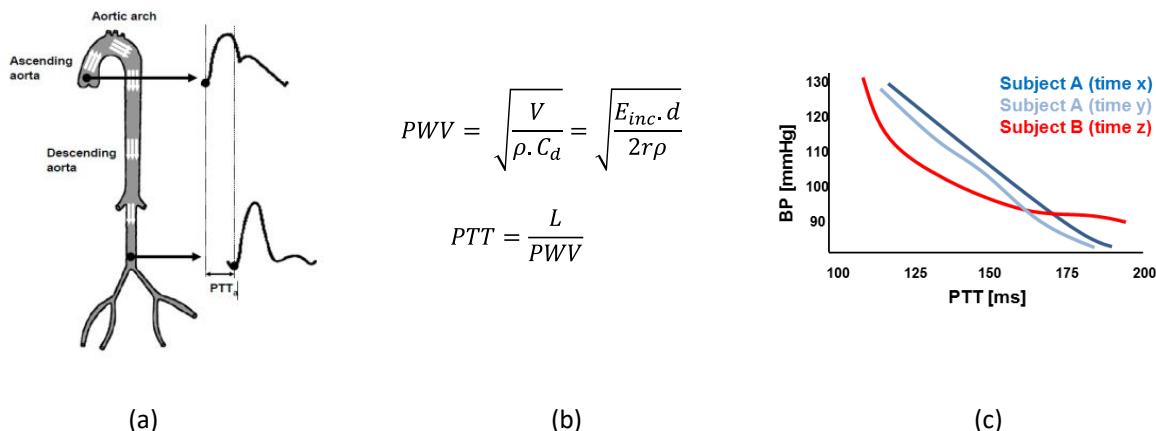


Figure 7-28 (a) Diagram to illustrate pulse transit time. (b) Equations for pulse wave velocity and pulse transit time. (c) Measured PTT versus blood pressure for various time and different test person.

In Figure 7-28 (a) a diagram is shown of the aorta where a blood pressure pulse is measured at two sites separated by a distance L. The pulse transit time (PTT) is defined as the time difference between the foot of the two pressure pulses. The pulse wave velocity (PWV) can be calculated easily from PTT when the distance is known. The pulse wave velocity (Figure 7-28 (b)) depends on the blood volume and elastic compliance of the artery (the Bramwell-Hill equation) which depends on the incremental Young modulus E_{inc} (Moens-Korteweg equation). Compliance is a function of the arterial pressure (see Figure 7-2). When the relation between PWV and arterial pressure is known (calibration method is a challenge because multiple blood pressure values are needed) the value of PTT can be used to estimate arterial blood pressure. Measurements of PTT for two subjects A and B and different times of the day (subject A) are shown in Figure 7-28 (c). There is a large person to person variation in the curves and there is a significant time dependence of PTT. The vasoconstrictor tone of the muscular arteries varies over time. As a result, the volume and elastic compliance of the arteries and thus PTT

varies with time. In practice the calibration curve must be determined regularly. Blood pressure and PTT measurements must be available over a large range of transmural pressures while other parameters remain fixed. This is not straightforward and there is no good calibration procedure yet. Note that the value of aortic PWV does have clinical relevance. It is a strong predictor for the probability of cardiovascular events. It is used for screening purposes.

PTT and PWV of the aorta can be measured using non-invasive methods. Two tonometry sensors are placed at the femoral artery and carotid. An ECG measurement is used to subtract the transit time from the heart to the carotid artery from the carotid-femoral artery PTT. This measurement requires a trained clinician and is most frequently done in the hospital during examinations.

For home monitoring or for lifestyle devices a simpler embodiment is needed. The pulse arrival method (PAT) has been proposed for smart watches and simple consumer devices. This measurement is illustrated in Figure 7-29 (a). The difference between PTT and PAT is indicated in this figure. PTT is the transit time between two sites in a large artery, it is a true measurement of the PWV in that arterial segment. PAT is the time difference between the R-peak of an ECG signal and the foot of a peripheral arterial pressure pulse wave. PAT is the sum of PTT (the travel time of the arterial pulse) and the pre-ejection period (PEP). PEP is roughly equal to the duration of the isovolumic contraction time. It is the time difference between the start of contraction (R wave of the ECG) and start of ejection of blood into the aorta. The average duration of PEP is approximately 75 milliseconds and depends on preload, afterload, contractility and breathing. PEP increases for some chronic heart diseases. PTT is of the order of 120-150 ms, the PEP period is about 30% of PAT. Hence interpretation of the measured PAT in terms of blood pressure changes is more complex due to the effects of PEP with blood pressure.

PAT requires an ECG signal and a surrogate for an arterial pressure signal at a peripheral site. Often a ppg signal from the wrist or finger is used. For home monitoring ECG electrodes are not ideal (long use time causes skin irritation, not compatible with shower etc.) and other embodiments have been proposed. A recent implementation in a smart watch is shown in Figure 7-29 (b). The ECG signal is measured between the two hands when the watch is touched by a finger from the other hand. The ppg signal is measured with a green LED ppg signal at the wrist-watch interface. Another implementation was the Xanadu device which uses the forehead for a ppg measurement and the finger and head for the ECG signal. Such devices are not used in a clinical environment yet, the deviation from reference auscultatory measurements is too large. Development of the Xanadu device has stopped. However, many of the manufacturers are aiming to develop continuous NIBP techniques based on pulse wave velocity.

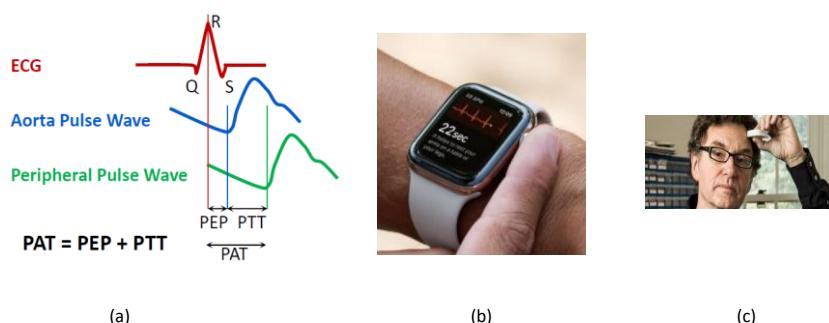


Figure 7-29 Diagram that shows the difference between PAT and PTT, PEP is the pre-ejection period. (b) Smartwatch implementation of ECG and possibly PAT (c) Xanadu device.

There are many challenges that need to be overcome before these techniques can reliably detect blood pressure or even pressure changes. PWV is a measure of compliance only, not of BP, factors that influence compliance will influence the measurement (or give a rise to the need of recalibration). There are more unknowns than measurements for the calibration (1 PTT to infer SBP, DBP and MAP). The determination of the location of the foot of the pulse or other location in the leading edge of the pulse is a challenge, the sensitivity of PAT is about 1ms/mmHg. Calibration is subject and time dependent and is very difficult. Finally, the effect of posture and hand position needs to be corrected (effect of hydrostatic pressure on compliance and PWV).

There are other methods that analyze the ppg waveform, detect reflections and infer blood pressure from this analysis. No reliable measurement has been shown up till now. They are not discussed in this chapter. The measurement of blood flow is discussed in the following section.

7.8 Blood Flow and Cardiac Output

Maintaining oxygen delivery to tissues is the first-priority during critical care. Oxygen delivery (DO_2) is defined as the volume of O_2 gas delivered per minute to the systemic vascular bed. It is the product of cardiac output (CO , l/min) and the arterial oxygen concentration (CaO_2 : ml O_2 gas per liter blood). Cardiac output is the blood volume per minute that is pumped by the heart into the aorta. It is between 4l/min and 6 l/min during rest. The equations that are relevant for oxygen delivery are listed below.

	$DO_2 = CO \times CaO_2$	7.8-1
	$CaO_2 = 13.9 \times [Hb] \times SaO_2 + PaO_2 \times 0.003$	7.8-2

$[Hb]$ is the hemoglobin concentration per liter blood (gr/l) and S_aO_2 is the arterial saturation of hemoglobin with oxygen. P_aO_2 is the partial pressure of oxygen in arterial blood (mmHg). The hemoglobin concentration, oxygen saturation and partial oxygen pressure can be measured in an arterial blood sample. The oxygen delivery should be larger than the oxygen consumption in the tissues. The normal oxygen consumption is between 180 ml/min and 250 ml/min. For a healthy person the oxygen delivery is larger than 1000 ml/min and there is a large overcapacity. During critical care conditions this may not be the case anymore and the clinician must apply therapy to the patient to prevent tissue hypoxemia and acidosis. Measurement of cardiac output in such conditions is a must. Cardiac output is the best indicator of the status of the circulation but is difficult to measure. A diagram of the options to measure cardiac output is shown in Figure 7-30. An indication of the type of measurement and the gold standards are indicated.

The measurements can be divided in two groups. Measurements that measure direct beat to beat flow and indirect cardiac output measurements. Two measurements that measure direct flow are the continuous and pulsed ultrasound technique. The indirect methods are used more frequently in clinical use cases and are discussed after the direct flow techniques.

7.8.1 Cardiac Output - Direct Flow

The ribcage shields the heart and aorta from direct ultrasound imaging. Locations at the chest surface that are used for cardiac imaging are not optimal for cardiac output monitoring. The esophagus is a minimally invasive access point for imaging of the heart and large arteries and for monitoring flow velocity. The aortic flow can be measured using a Doppler ultrasound technique. There are two options, continuous ultrasound (flow velocity) and pulsed ultrasound (imaging plus flow velocity).

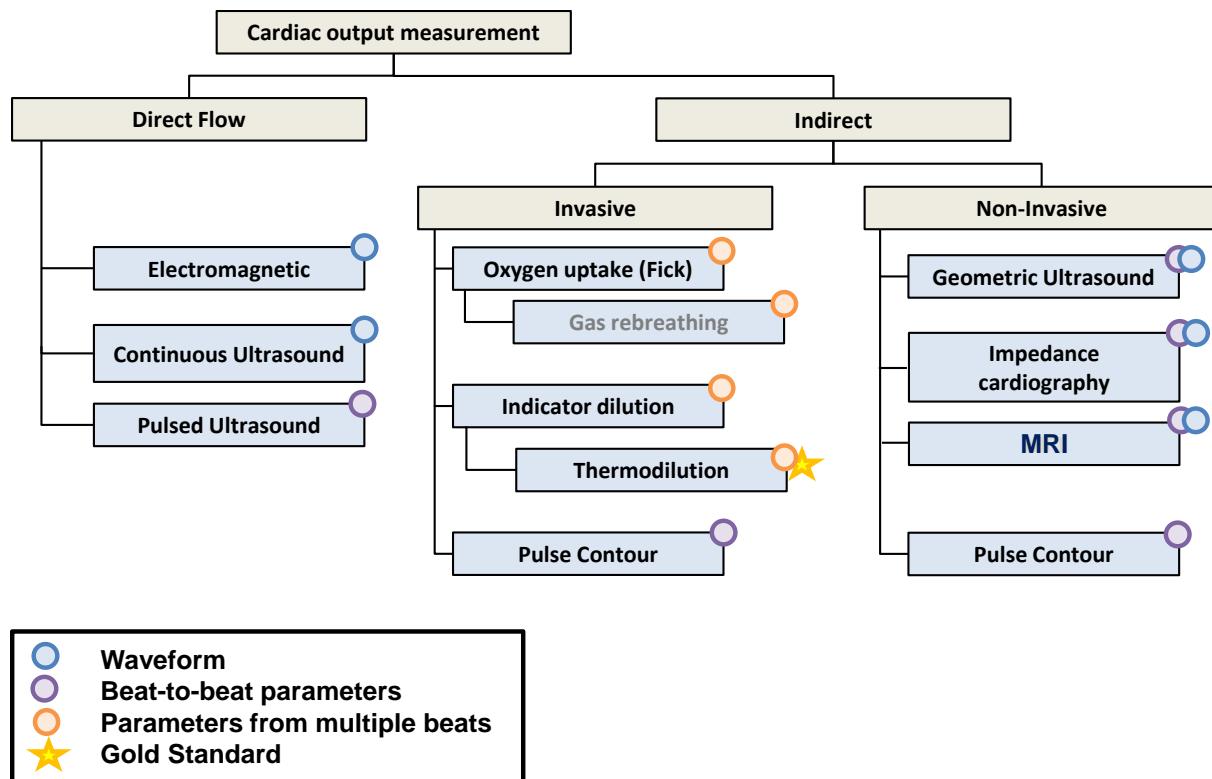


Figure 7-30 Diagram of the different invasive and non-invasive measurements for cardiac output. Gold status measurements are indicated.

7.8.2 Continuous ultrasound

A diagram of a commercial catheter system to monitor cardiac output via the esophagus is shown in Figure 7-31. The flow velocity in the descending aorta is measured with a continuous-ultrasound Doppler transducer. A disposable catheter with an ultrasound probe is inserted into the esophagus via the oral cavity (nasal cavity is also possible). The probe is adjusted and aligned to obtain the optimum signal. The probe has an angle of 45 degrees with the aorta to minimize errors due to angle dependence in the Doppler measurement.

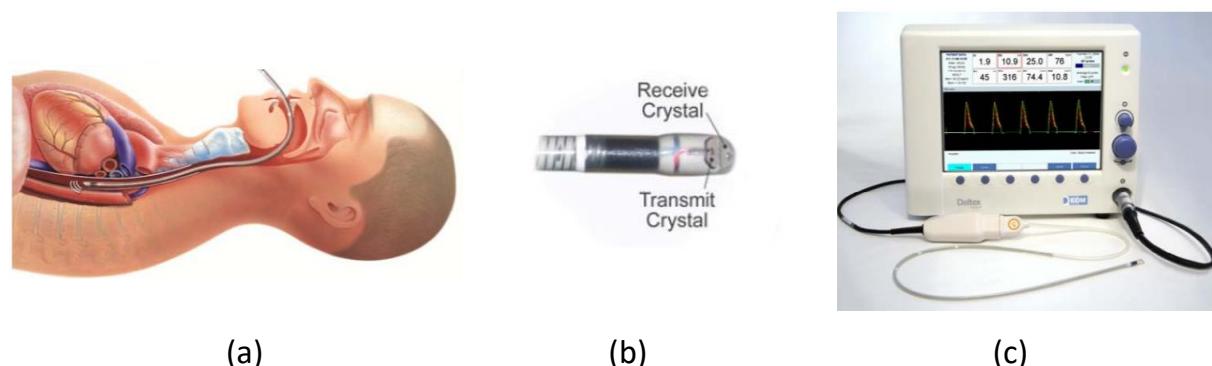


Figure 7-31 Deltex esophageal continuous ultrasound system for measurement of beat to beat stroke volume and cardiac output. (a) Schematic drawing of probe inserted via the oral cavity, (b) ultrasound probe with separate send and receive crystals, (c) monitor to display velocity waveform and parameters.

The reflected beam is frequency shifted due to the Doppler Effect; the frequency shift is converted in a beat-to-beat flow velocity versus time via the well-known equation:

$$v = \frac{c \cdot f_d}{2 f_t \cos \theta}$$

7.8-3

Where c is the velocity of sound, f_d is the Doppler frequency shift, f_t is the transmitted frequency and θ is the angle between the flow direction and the probe transducer. The flow in the descending aorta is approximately 70% of the total cardiac output. The diameter of the aorta is not known and cannot be measured with this system. A proprietary method is used to estimate and convert flow velocity into stroke volume and cardiac output. This method is based on a "calibration" of the measurement using data from many studies that used both this system and a gold standard thermodilution method. Furthermore, population data is used to convert flow velocity into cardiac output. Hence it is not an absolute measurement of the cardiac output but the advantage is that it offers a beat-to-beat signal that is proportional to cardiac output, has a good precision and the shape of the signal has "analog" information that can also be of great use to the clinician. This is illustrated in Figure 7-32 for a patient during surgery which became hypovolemic and was having a low cardiac output but also had high heart rate and anomalies in the flow pattern (short duration pulses, small stroke volume and lower peak velocity).

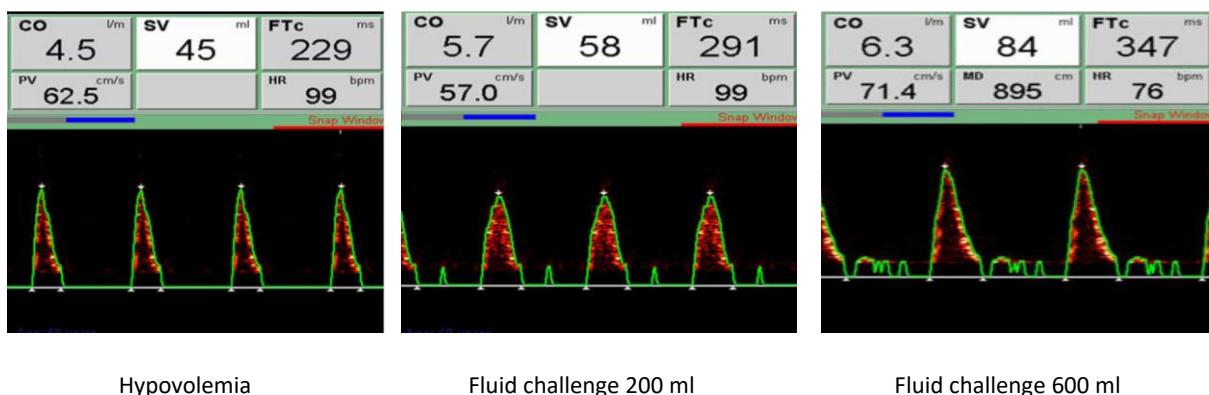


Figure 7-32 Velocity spectrograms for a patient with hypovolemia, after a fluid challenge of 200 ml and after a challenge of 600 ml.

7.8.3 Pulsed ultrasound TEE Ultrasound

Pulsed ultrasound combines the image and position dependent ultrasound Doppler flow velocity. The total flow can be calculated when the lumen diameter and flow velocity profile has been measured. Trans Esophageal Echo (TEE) is a high-end ultra-sound transducer technology for diagnostic use cases (heart function, valve diseases) and for use during complex cardiac surgery. This transducer offers the highest quality 2D and 3D imaging and Doppler ultrasound of the heart, heart valves and large arteries around the heart. The transducer is shown in Figure 7-33. It is a combination of an endoscope and a top-quality 3D pulsed ultrasound transducer. The transducer device cost is very high (> 50k€). The device can be re-used for other patients after sterilization in an autoclave. The skill and training level of the operator must be very high. High-end ultrasound machines are needed. The transducer is not suited for routine use, it is mostly used during cardiac surgery in specialized hospitals. The unique combination of features (real-time 3D imaging and color Doppler flow) is very attractive but for an absolute measurement of the cardiac output the device is less suited. Small errors in the determination of the aorta dimensions and errors in angles with respect to the flow direction lead to sizeable errors in the cardiac output. For more accurate cardiac output measurements the indirect measurements are more suited.



Figure 7-33 Philips TEE transducer and 2D color Doppler of the aortic valve of the heart.

7.9 Indirect Cardiac Output Measurements

The most accurate measurements of cardiac output are done with indirect measurements. The first measurement of cardiac output was proposed by Fick in the 19th century was based on the uptake of oxygen in the lung. More practical measurements are based on indicator dilution methods. These two types of measurements are invasive and complex and are described in the following sections.

7.9.1 Fick Principle

The Fick principle is illustrated with the diagram, shown in Figure 7-34.

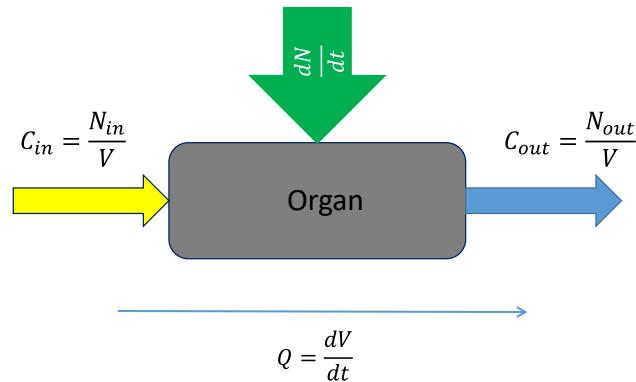


Figure 7-34 Illustration of Fick's principle. C_{in} and C_{out} are the concentrations of a marker. Q is the volumetric flow.

An amount $\frac{dN}{dt}$ of a marker is taken up by an organ per unit time. The organ receives blood with marker concentration C_{in} and blood leaves the organ with a concentration C_{out} . The volumetric blood flow Q through the organ is the volume of blood that flows per unit time through the organ. The amount of marker that flows out per unit time is:

	$\frac{dN_{out}}{dt} = \frac{dN_{in}}{dt} + \frac{dN}{dt}$	7.9-1
--	--	-------

The amount of marker that enters or leaves the organ per unit time is equal to the product of volumetric flow Q and concentration C . Equation 7.9-1 can then be rewritten as:

	$\frac{dN}{dt} = (C_{out} - C_{in}) \cdot Q$	7.9-2
--	--	-------

The uptake of the marker by the organ is proportional to the difference of the outflow and inflow concentrations and the volumetric flow. When the uptake and the two concentrations can be measured the flow Q can be calculated. According to this principle the cardiac output CO can be measured using oxygen as an indicator. The uptake of oxygen per minute in the lung (\dot{V}_{O_2}) can be

measured with a spirometer. The concentrations of oxygen in the mixed venous blood in the pulmonary artery (C_{PA}) and a large artery of the systemic circulation (C_{art}) can be measured in blood samples. This is illustrated in the diagram of Figure 7-35 (a). This gives:

$$CO = \frac{\dot{V}_{O_2}}{C_{art} - C_{PA}}$$

7.9-3

The oxygen uptake can be measured with a spirometer, the concentrations of oxygen in the pulmonary vein and pulmonary artery require blood samples (see Figure 7-35 (b)). The Fick technique requires invasive sampling of arterial and mixed venous blood. The oxygen concentration of oxygenated blood is measured in the large systemic arteries, no oxygen is consumed yet. The radial artery and femoral artery are often used. For the venous side mixed venous blood in the pulmonary artery is needed as oxygen concentrations in the inferior and superior vena cava are different. The mixing occurs in the right heart. The best location to measure the oxygen concentration of deoxygenated blood is in the pulmonary artery. Sampling of blood requires the use of a pulmonary artery catheter that must pass the right heart before it reaches the pulmonary artery. This is a difficult procedure that poses a high risk to the patient.

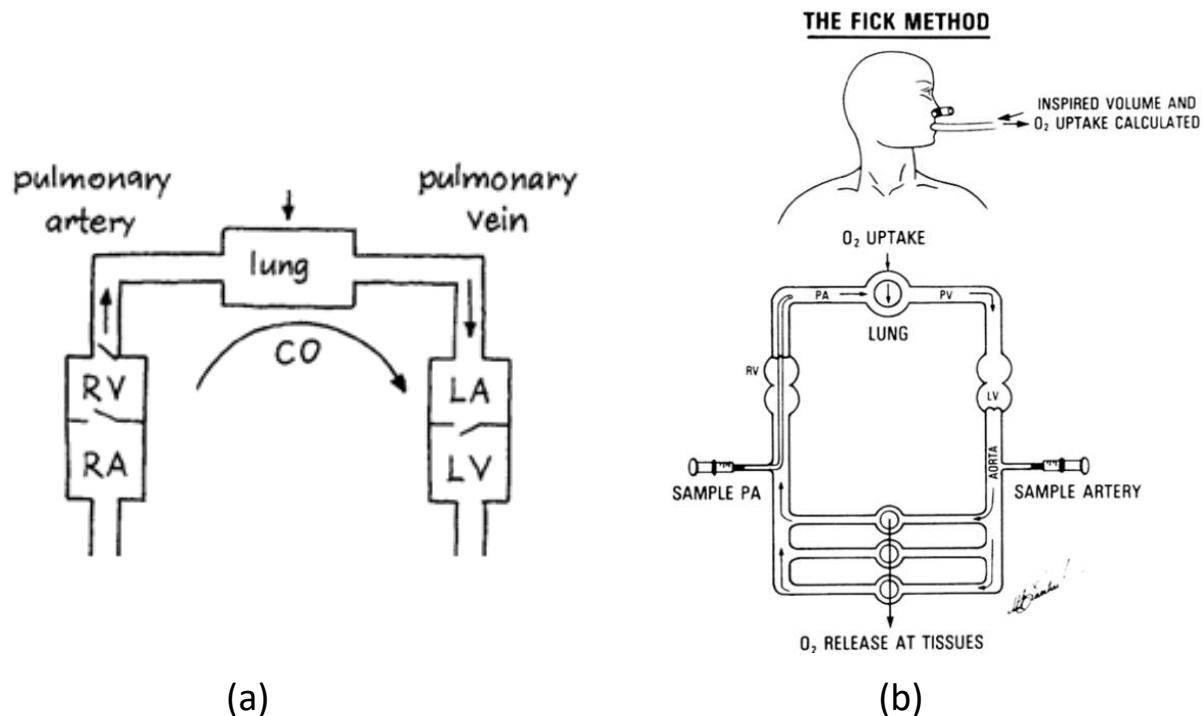


Figure 7-35 Diagram that shows the use of Fick's principle (a) and practical implementation (b).

Measurement of the cardiac output using the Fick principle and oxygen as a marker is difficult, risky and time consuming. It may take minutes before the measurement of the uptake of oxygen is available. Blood samples are submitted to the lab for analysis. The Fick method fails when there is a considerable shunt flow (blood flow to parts of the lung that are not ventilated well) through the lungs. In summary the method is complex, time consuming and sensitive to errors. Therefore, the Fick method is rarely used for cardiac output measurements. The method can be used to measure flow through a specific organ. For instance, renal flow is measured with a variant of this method. In short there is a need for more efficient and more accurate measurement, the indicator dilution

method was proposed and is now considered to be the gold standard for cardiac output measurement. It is described next.

7.9.2 Indicator Dilution Method

The principle is illustrated with the diagrams shown in Figure 7-36. A known mass of an indicator (for example a dye) is added to a fluid with unknown volume B. When after mixing the concentration of the indicator is measured, the fluid volume B can be calculated. A measurement of the concentration is often easier than a direct volume or flow measurement. This is an example of a measurement of a parameter that is difficult to measure and is replaced by a measurement of a parameter that is easier to measure.

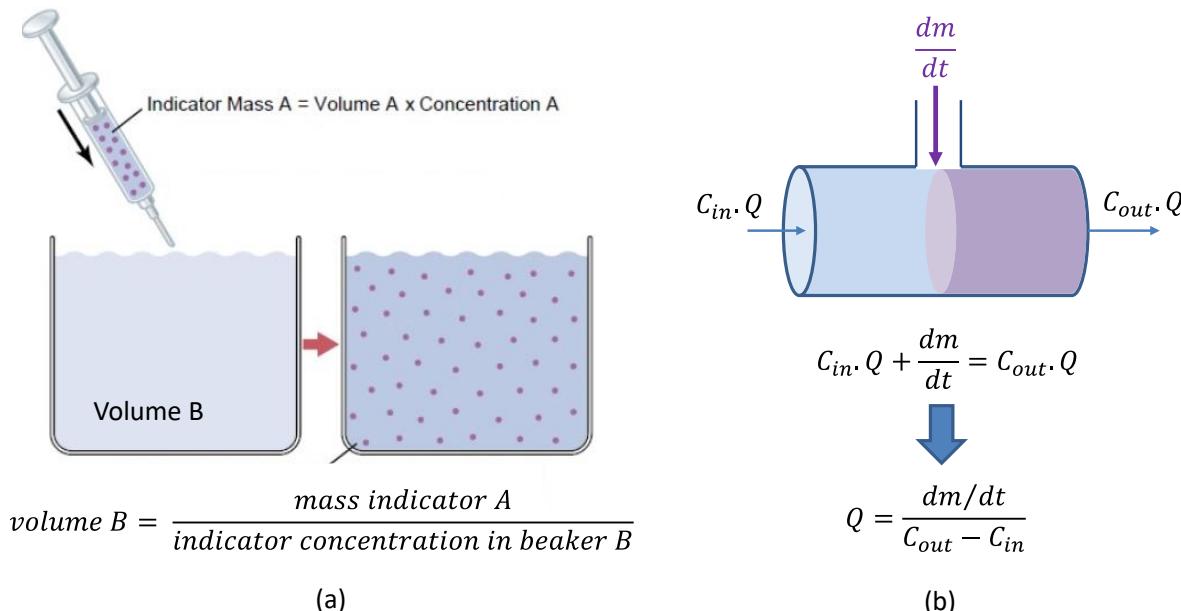


Figure 7-36 Injection of a known mass of an indicator A in to an unknown fluid volume B allows measurement of fluid volume B after a measurement of the indicator concentration (a). (b) Continuous injection of an indicator into a tube and downstream measurement of the indicator concentration allows measurement of the volumetric flow Q.

To measure flow continuously in a tube there is a continuous infusion of an indicator, and the downstream concentration is measured. It is possible to calculate the flow Q using equation 7.9-3 (see Figure 7-36 (b)). This is a variant of the Fick method; the method and mathematics are identical. This measurement is not practical as large amounts of indicator must be used and this indicator will return to the injection site after a full loop in the circulation.

This has been solved in the rapid injection indicator solution which is illustrated in Figure 7-37. The drawback is that the measurement is intermittent.

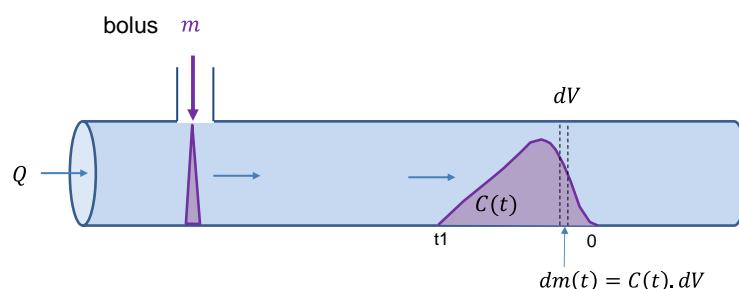


Figure 7-37 Diagram to illustrate the rapid injection indicator dilution method to measure volumetric flow rate Q.

A known bolus (i.e. a fast injection of a known indicator mass m) is injected into the blood stream and the indicator concentration is measured downstream. The indicator is a substance that can be measured easily and with high accuracy. It is non-toxic and should disappear quickly from the blood volume after the measurement has been done. The path from the injection site to the measurement site should have no branches which add blood without indicator to the artery before the measurement site. It should also not be absorbed by the vessel wall. Many indicators have been used or are still in use including radioactive tracers, chemical substances, dyes, and heat (thermodilution, cold saline fluid).

Assume that a small bolus of an indicator (known mass m) is injected in the vena cava superior near the right atrium and is well mixed during the flow through the right heart. The indicator concentration is measured in a large artery of the systemic circulation or pulmonary artery (blood ejected by the right heart flows through the pulmonary artery and lungs via left heart to the artery). The concentration of the indicator in arterial blood is the same in the large arteries, the measurement of the indicator concentration can be done at the most convenient site. The mass $dm(t)$ of the indicator solution in the volume dV at the measurement site at time t is equal to $dm(t) = C(t) * dV$. The indicator mass per unit time that passes the measurement site between times t and $t + dt$ is given by:

$$\frac{dm}{dt} = C(t) \frac{dV}{dt} = C(t) \cdot Q \quad 7.9-4$$

Integration over time of equation 7.9-4 gives:

$$m = \int_0^{t_1} C(t) \cdot Q \, dt \quad 7.9-5$$

Assuming that Q is constant we obtain an expression that relates Q to the bolus indicator mass m and the time integral of the concentration (Stewart-Hamilton equation):

$$Q = \frac{m}{\int_0^{t_1} C(t) \, dt} \quad 7.9-6$$

The time average of the flow Q is measured, and in this case, it is equal to the cardiac output CO . Two concentration profiles for dye (a) and thermodilution (b) are shown in Figure 7-38.

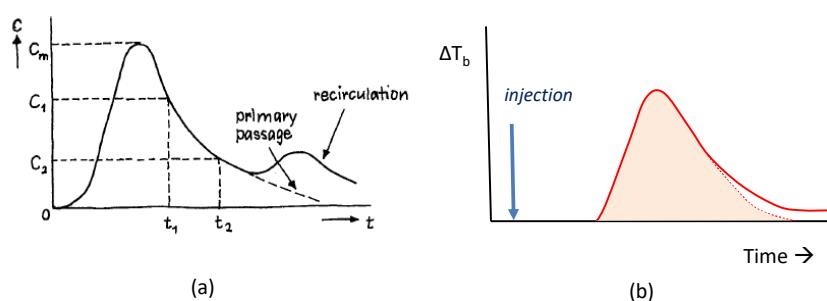


Figure 7-38 Dye concentration profile and temperature profile in thermodilution.

The tail of the primary distribution follows an exponential curve. The dye remains longer in the circulation and a signal due a recirculation component is added at larger times. This undesirable

signal is removed by an (exponential) extrapolation of the primary signal and integrating the corrected extrapolated signal (dashed line in Figure 7-38 (a)). The recirculation component is much smaller or even absent in the thermodilution technique. The dye indicator technique for cardiac output requires continuous withdrawal and sampling of arterial blood and optical measurements outside the body.

The thermodilution technique with a single pulmonary artery catheter (PAC, Swan-Ganz) does not require blood sampling for cardiac output measurement, it is the best location for the cardiac output measurement, and it has become the gold standard technique for cardiac output measurement. A diagram of the PAC thermodilution catheter after insertion in the pulmonary artery is shown in Figure 7-39. The catheter is inserted into the internal jugular vein or subclavian vein and shifted towards the right atrium. The balloon at the tip is then inflated and the blood flow drags the balloon and catheter via the right atrium and right ventricle into the pulmonary artery. The tip of the catheter should be at the position of the branching of the pulmonary artery. The catheter has a port (1) in the right atrium for injection of the cold saline or cold dextran solution into the right atrium. The cold solution is mixed with warm blood during its passage in the right heart and reaches the thermistor (2) of the pulmonary artery catheter where the temperature is sensed. There is a port (3) for the balloon inflation and a fourth port for blood sampling (mixed venous blood) and pressure measurement of the pulmonary artery. The absolute value of the change in temperature is plotted in Figure 7-38 (b).

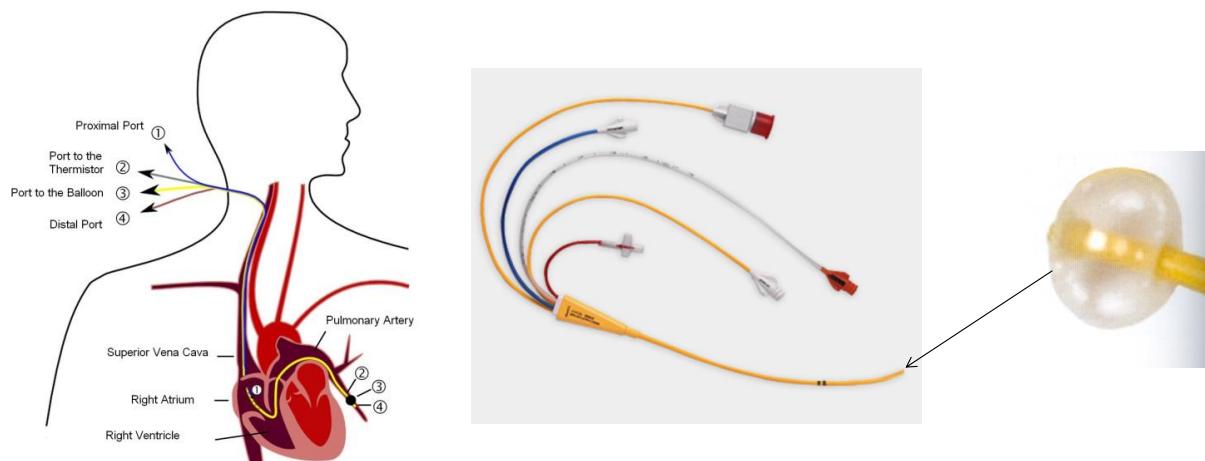


Figure 7-39 Pulmonary artery catheter for thermodilution cardiac output measurements. A balloon at the catheter tip can be inflated when inserted in the right atrium and drags the catheter through the heart into the pulmonary artery.

The cardiac output is calculated using the modified Stewart-Hamilton equation shown below.

$$CO = \frac{V_i \cdot (T_b - T_i) \cdot K}{\int_0^t \Delta T_b dt}$$

7.9-7

Here V_i is volume of the injected cold indicator solution, T_i is the temperature of the indicator solution, T_b is the blood temperature, K is a correction factor for the specific weight and specific heat of the indicator and blood, ΔT_b is the change in blood temperature measured at the thermistor site. The clinical practice is to perform three or more measurements in rapid succession and to take the average of the results which agree closely.

Following the introduction of the pulmonary artery catheter (PAC) into clinical practice, the single bolus thermodilution measurement of cardiac output has been widely accepted as the “clinical standard” for advanced hemodynamic monitoring. It is at present the clinical gold-standard against which new technologies are validated and compared.

The technique should be repeated many times at short time intervals. Furthermore, the PAC catheter allows blood sampling of mixed venous blood and enables the measurement of pulmonary artery pressures. The so-called wedge pressure can be measured, this is the pressure measured by wedging a pulmonary catheter with an inflated balloon into a small pulmonary arterial branch. It estimates the left atrial pressure. The flow, pressure and blood gas parameters are of great importance for therapy during critical care.

There several disadvantages of the technique. The measurement requires a highly skilled operator, the fluid injection speed and injected volume are critical parameters, rewarming of the solution by wall material, time of injection, dead space volumes in the catheter are factors that need to be eliminated. The importance of the time of injection relative to the lung pressure wave during mechanical lung ventilation and resulting time dependence of cardiac output is illustrated in Figure 7-40. The measurement of cardiac output using a Swan-Ganz catheter is plotted versus time during a positive pressure ventilation cycle. During inspiration the cardiac output is severely reduced (40%) followed by an overshoot (20%). This illustrates that for the highest accuracy cardiac output data an average of multiple measurements during a ventilation cycle are needed. The timing of the measurement is important.

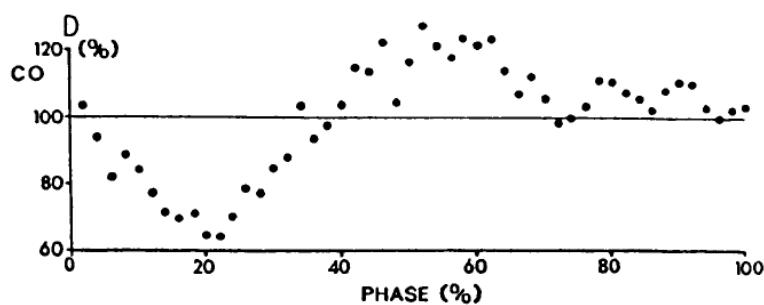


Fig. 4. De metingen van Figuur 3 geordend naar de beademingsfase.
(Jansen, proefschrift Rotterdam, 1988, p42)

Figure 7-40 Measurements of cardiac output as function of the phase in the ventilation cycle.

The main disadvantages of the PAC catheter are the morbidity and mortality due to its use. Placing a pulmonary artery catheter is highly invasive, can induce infections, arrhythmias and even ventricular fibrillation, ruptures and hemorrhage. The use of a PAC catheter poses a risk to the patient which does not seem to be balanced by an improved outcome. The discussion about its use is still ongoing.

Nevertheless, the measurement of the cardiac output using a swan-Ganz catheter is the gold standard method against which other techniques have to be validated and compared. There is no better solution at the bedside at this moment in time. There is a high need for beat-to-beat measurements of stroke volume and cardiac output, this data cannot be provided by a thermodilution method. Techniques that are capable of continuous cardiac output measurement are discussed in the following section.

7.9.3 Model Based Methods – Pulse Contour Methods

A beat-to-beat measurement of the stroke volume and cardiac output produced by the heart would provide the best possible information to the clinician about the heart pump and circulation functions. In principle a TEE ultrasound probe could provide such information, but lack of experienced operators, accuracy and cost are preventing widespread use. An alternative method was proposed by Frank in the 19th century. It is based on the Windkessel model of the circulation. The method is illustrated with the diagram shown in Figure 7-41. It is the pulse contour method.

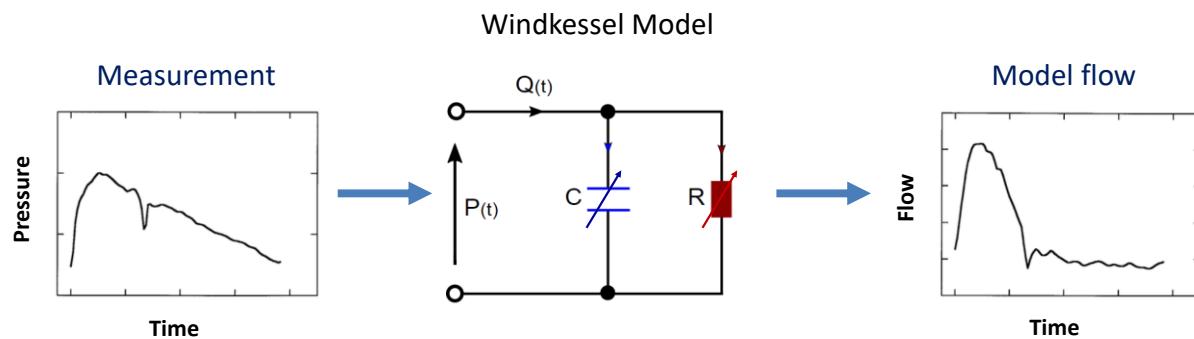


Figure 7-41 two-element Windkessel model with measured pressure as input parameter. The flow out of the heart can be calculated when the values of the systemic resistance and arterial compliance are known.

The arterial pressure generated by the heart can be measured more easily than the stroke volume. Using a Windkessel model of the output load of the heart the flow at the output of the heart and the stroke volume can be calculated when the values of the arterial compliance C and systemic resistance R are known. Unfortunately, the values of R and C cannot be measured directly.

Pulse contour methods faces many challenges for the accuracy and precision.

1. The compliance is non-linear, and this complicates the analysis. Furthermore, the compliance and its non-linearity are patient specific.
2. Systemic resistance is not known.
3. Reflection of pressure waves from the periphery add to the forward wave. These pressure waves must be separated before the stroke volume can be calculated.
4. Pressure measurement with a fluid filled catheter may cause errors due to non-optimal damping settings.
5. The aortic outflow is more complex during the systolic period than assumed in the model.
6. A regular heartbeat is needed.
7. Compliance and systemic resistance may vary rapidly, errors in parameters directly affect the calculated cardiac output.
8. Methods are not accurate during critical conditions with severe vasoconstriction.

Several methods have been developed to solve to obtain values for the unknown R and C . In a first method population averages (gender, age) of the non-linear arterial compliance are used. The systemic resistance is determined in an iterative manner. In the second method a calibration step of the cardiac output precedes the measurement and is used to determine the values of the parameters of the Windkessel model. These methods are briefly described below.

7.9.3.1 Pulse Contour Method – Population Based Compliance

When the non-linear arterial compliance and central blood pressure pulse is available the stroke volume output can be computed. The values for the non-linear compliance were obtained from measured volume-pressure relation and elastic compliance of human aortas extracted from human

cadavers. These data were fitted with a simple analytic function (arc tangent). The fit parameters of this function correlated with age and gender. For a specific patient the non-linear arterial compliance is estimated by using the age and gender and thereafter the stroke volume is computed according to the method developed by Wesseling. The method is illustrated in Figure 7-42.

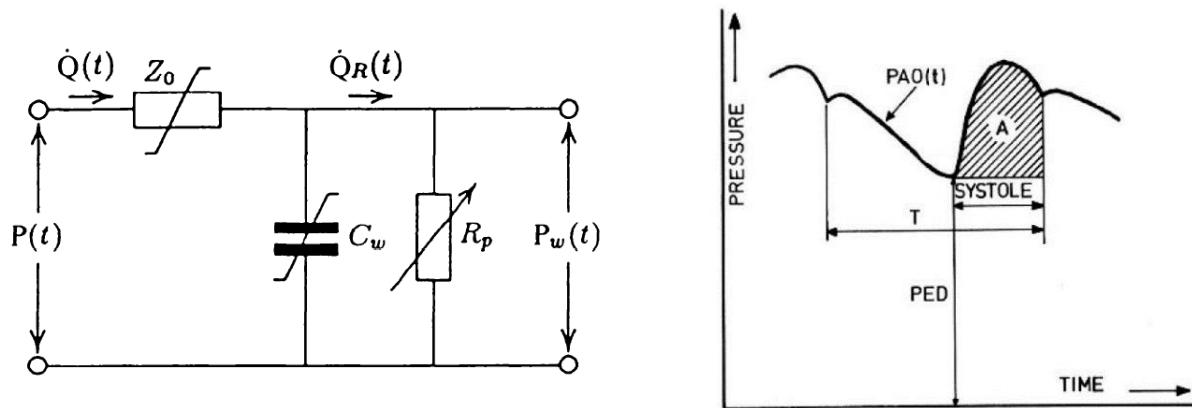


Figure 7-42 three element Windkessel model and arterial waveform. The dashed area A is used to calculate the stroke volume.

The arterial pressure is measured (right figure) and is the input pressure of a three-element Windkessel model. The characteristic impedance Z_0 and arterial compliance function C_w are estimated using the age and gender of the subject. A default value for the systemic resistance R_p is used for the first stroke. The first stroke volume is computed using the systolic part of the pressure pulse (i.e. hatched area A of the pressure curve) and first value of the systemic resistance. The value of the systemic resistance is evaluated beat by beat and is adjusted in an iterative manner such that the calculated pressure pulse approaches the measured one, it converges rapidly to a steady state value that is used to calculate the stroke volume. The heart rate is easily measured, the cardiac output is equal to the product of heart rate and stroke volume. The method described above was developed for the Finapress device which measures the blood pressure non-invasively using the volume-clamp technique. In this way a non-invasive beat-to-beat estimate of the cardiac output was made possible.

Unfortunately, the accuracy of this method is questionable. Many studies have been performed with mixed results. The cardiac output is measured most often in unstable patients where peripheral resistance and arterial compliance can change rapidly. This can lead to large errors in the estimation of the cardiac output. Furthermore, the use of a population average for the arterial compliance introduces large errors, there is a large person to person variation in anatomic structure and compliance values of the arterial system. Therefore, by other device suppliers a calibration method was used to obtain values for the non-linear compliance and to personalize the continuous cardiac output measurement. These methods are used widely and are described below.

7.9.3.2 Calibrated Pulse Contour Method – PiCCO

The PiCCO (Pulse Contour Cardiac Output) solution for beat to beat cardiac output are described below. PiCCO uses a combination of two techniques for monitoring of cardiac output and volume status (see Figure 7-43). A transpulmonary thermodilution measurement is used for calibration of the stroke volume and determination of volumetric parameters. A pulse contour method is used for the continuous cardiac output. A central venous line is placed via the subclavian or jugular vein. An arterial line with an integrated thermistor is inserted via the femoral artery and is placed in the thoracic aorta. It measures both pressure and temperature. A cold saline fluid is injected via a central venous line into the right atrium. The indicator flows through the right heart and is mixed

with blood. Thereafter the indicator flows through the pulmonary circulation and is pumped via the left heart into the aorta. The blood temperature is sensed in the aorta. The calibration cardiac output is determined by a thermodilution method. The transpulmonary dilution method is less accurate than the gold standard pulmonary artery catheter method. An advantage of transpulmonary thermodilution is that it is independent from breathing or ventilator cycles. Additionally, because the indicator passes through the heart and lungs, this allows the determination of intravascular and extravascular fluid volumes inside the chest area, the preload volume and lung water". The aortic pressure is sensed in the aorta close to the aortic valve. The cardiac output is determined using a two-element Windkessel model (see Figure 7-44 (a)).

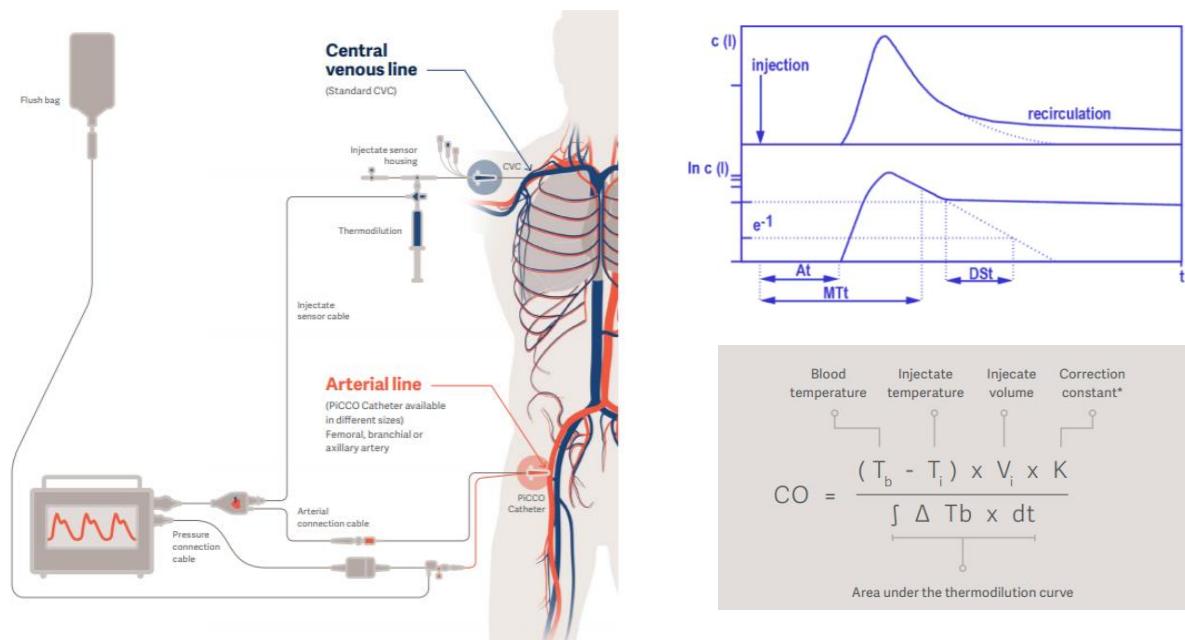


Figure 7-43 Diagram of the PiCCO system for measuring cardiac output. Temperature change measured in the aorta and parameters derived from the temperature change.

The stroke volume SV is determined by integration of the Windkessel model flow Q(t) due to pulse pressure $P_p(t)$ over the systolic ejection period (Figure 7-44 (b)). The initial transpulmonary dilution method is used to calibrate the parameters and a proprietary algorithm calculates the beat by beat cardiac output (see Figure 7-44 (b)). The compliance is determined from the calibration data for cardiac output and arterial pressure pulse. The systemic resistance R (or SVR) can be estimated from the relation $CO = (MAP - CVP)/R$ and subsequently determined by an iterative method. Frequent recalibration may be needed when the hemodynamic status is unstable.

The PiCCO technique is less reliable when the cardio-vascular patient status changes rapidly. Any deviation from the actual modeled elements translates directly into an error in the cardiac output. Both compliance and systemic resistance may vary rapidly. Furthermore, the heart rate must be regular, the method is not advised for patients with cardiac arrhythmias. Not all parameters can be measured simultaneously, and estimated values may not be accurate. A more reliable method to estimate the vascular status (compliance, resistance) is needed.

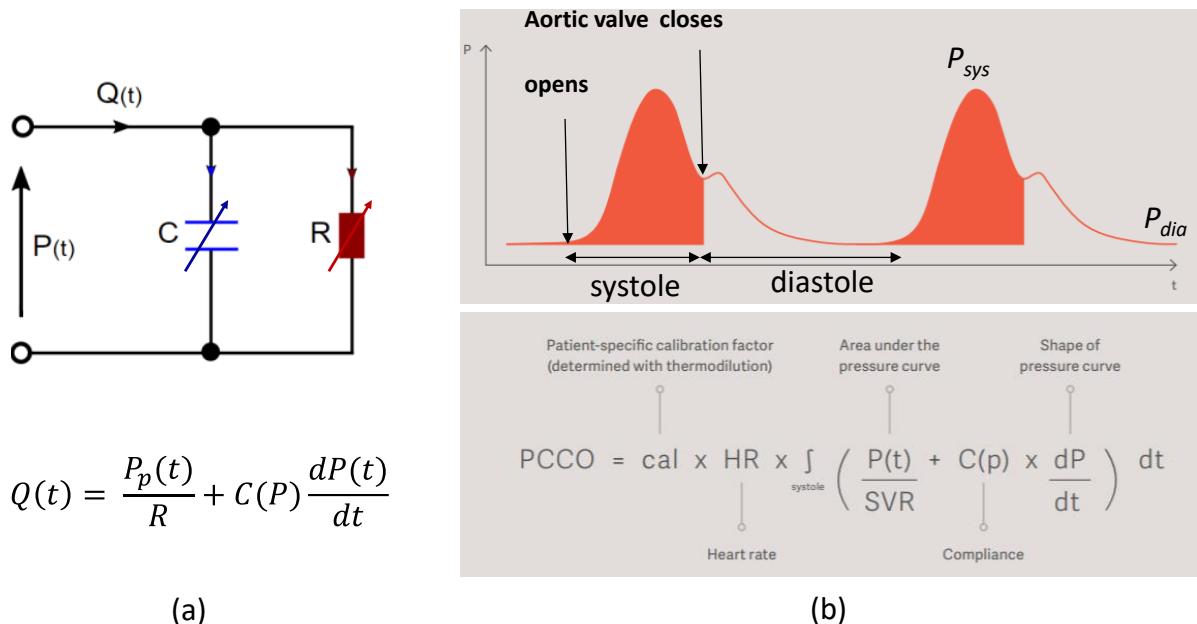


Figure 7-44 (a) Two element Windkessel model and equation for flow $Q(t)$. P_p is the pulse pressure. (b) Integration of systolic flow gives the stroke volume and cardiac output.

7.9.3.3 LidCO Cardiac Output Method

This technology uses an invasive or non-invasive arterial pressure pulse to estimate cardiac output. The pressure signal that is used can be a copy from the signal of a patient monitor. The arterial measurement site is not restricted. The basic steps of the method are shown in Figure 7-45.

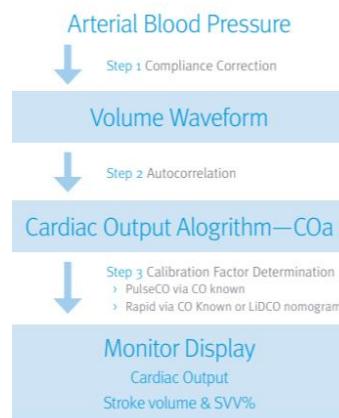


Figure 7-45 Steps in the LidCO cardiac output measurement.

The arterial blood pressure waveform is converted into an arterial volume change via the approximate equation:

$$\frac{\Delta V}{\Delta P_a} = \text{cal} \cdot V_{amax} \cdot e^{-k \cdot P_a}$$

7.9-8

Where ΔP_a is the measured blood pressure change, P_a is the arterial pressure and k is a fixed curve coefficient and the parameter V_{amax} (default value is 250 ml) is the maximum additional volume that can be added to the arterial tree. After the conversion from a pressure to a volume waveform, an autocorrelation algorithm of the waveform (20 second periods) provides the heart rate and a net

effective beat power factor (i.e. proportional to the root mean square of the volume pulsations). This power factor is proportional to the net effective stroke volume ejected into the aorta. Using the default parameters, a nominal stroke volume is obtained (SV_{nom}). The nominal stroke volume can be converted in an absolute stroke volume (SV_{cal}) after the calibration procedure. The parameter k is a constant, the calibration factor corrects and personalizes the V_{max} volume.

The cardiac output calibration is done with a proprietary Lithium ion dilution technique, an injection of an isotonic Lithium Chloride solution in a central or peripheral vein is used (see Figure 7-46). The Lithium ion concentration can be measured in any artery, a small volume of blood is continuously pumped into an ion-specific electrochemical sensor which is used for the concentration measurement. The calibration method is less invasive and safer than the pulmonary artery catheter.

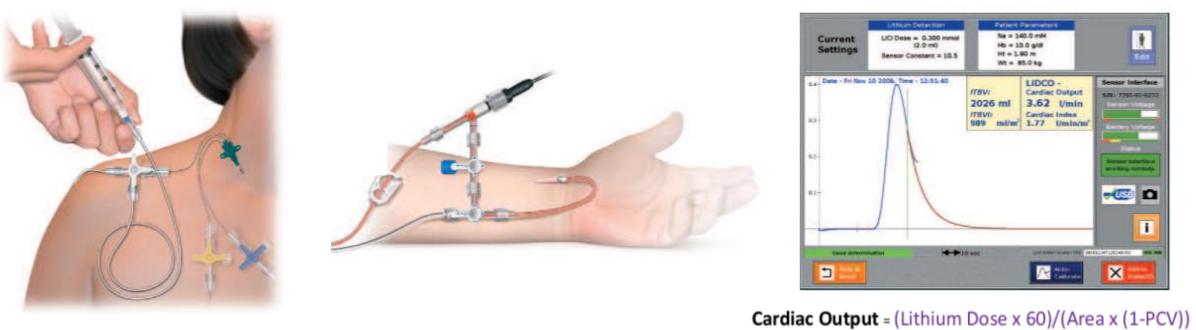


Figure 7-46 Calibration of the cardiac output by a lithium ion dilution method.

LidCO claims to have less error sources than the pulse contour method. Note that changes in arterial compliance directly affect the calibration. The method should not be used in cases of severe vasoconstriction and aortic valve diseases. The electrode is sensitive to other ions and interactions can occur (e.g. muscle relaxants). When the cardiac output change is large a new calibration run is needed.

When different methods for measurement of cardiac output are compared the differences are often large (> 20%). There is even uncertainty whether the thermodilution method with a Swan-Ganz catheter is more accurate and precise than the LidCO method. Fortunately, a new reference method is available, it is a variant of MRI (Magnetic Resonance Imaging). It is discussed in the next section.

7.10 Velocity Encoded Phase Contrast MRI

Velocity-encoded phase contrast Magnetic Resonance Imaging is an accurate technique to measure flow in large blood vessels. Tests in phantoms with known reference flows have shown very good agreement with the reference flow measurement (error in the order of 5%). The variability and precision are better than the gold standard reference technique. However, MRI imaging cannot be applied routinely during clinical care.

Velocity-encoded MRI is based on the detection of changes in the phase of a. These phase changes are proportional to the flow velocity when the particles move in a well-defined magnetic gradient field. Two images are recorded per time interval in the cardiac cycle. The first is the anatomical picture where the location, orientation and vessel dimensions can be determined. The second is a picture in which the intensity of each pixel is proportional to the flow velocity. Integration of the flow over the cross section of the blood vessel gives the total flow. This is illustrated in Figure 7-47.

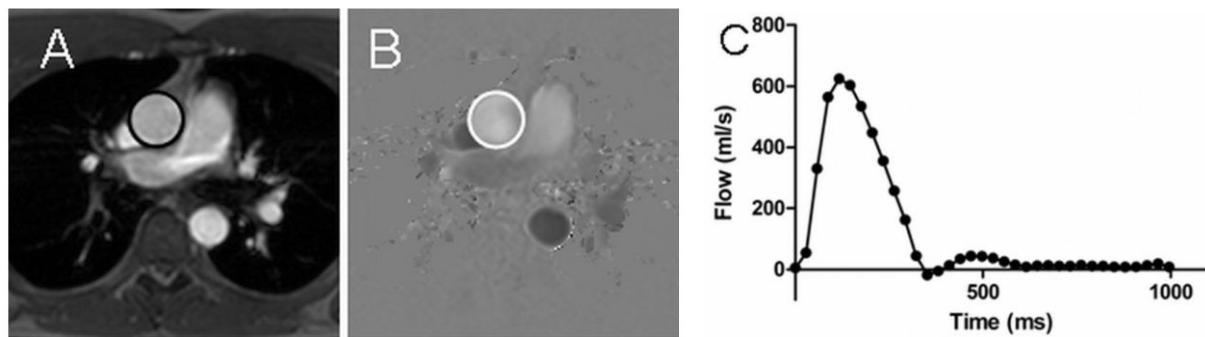


Figure 7-47: The modulus image (A) is used for anatomical delineation of the aorta (black circle) and measurement is performed in the corresponding phase image (B). Cardiac output can be calculated by quantifying stroke volume as the integral of the resulting flow curve (C) and multiplying with heart rate. (Carlson et al. Journal of Cardiovascular Magnetic Resonance 2012)

The patient must be transported to the MRI imaging system. Measurements at the bed side are not possible. Furthermore, the measurement is complex and time consuming and requires a skilled operator. The cost is high. MRI is a useful tool for research of cardiac output, it is not routinely used in clinical practice. MRI cardiac output is routinely measured during cardiac MRI examinations.

7.11 References

[Guyton] Guyton and Hall, Textbook of Medical physiology, Elsevier Health Sciences, ISBN 9781455770052.

[Webster] John G. Webster, Medical Instrumentation, Application and Design, John Wiley and Sons, fourth edition, ISBN 9780471676003.

[Westerhof] Snapshots of Hemodynamics, Nicolaas Westerhof, Nikolaos Stergiopoulos, Mark Noble, Berend E. Westerhof, Springer, ISBN 9783319919324, ebook.

[Blom] J.A. Blom, Monitoring of Respiration and Circulation, CRC Press LLC, ISBN 9780849320835

7.12 Questions

1. The "pulse contour method" has been proposed as a method to track changes in cardiac output. Describe the model that is the foundation of this method.
2. What are potential error sources in this method?
3. Describe Fick's method to measure cardiac output. Discuss the best locations to obtain blood samples.
4. Derive the formula that allows the calculation of the cardiac output from the observed concentration of an indicator passing a measurement site.
5. Show why recirculation need not be a problem in dye dilution cardiac output measurements.
6. At which anatomical sites can the blood pressure be measured with a catheter-manometer system? Assume that the catheter is long enough. Which sites are difficult or even impossible to reach with a catheter?
7. At which anatomical sites can the blood pressure be measured non-invasively? Which is the preferred site?
8. What is the "wedge pressure", how is it measured, and which information does it provide?
9. Discuss why the peripheral arterial pressure waveform differs from the pressure that can be measured at the root of the aorta.
10. What are a catheter-manometer system's natural frequency and damping ratio? What values should they have if the blood pressure is to be measured accurately up to 40 Hz?
11. Describe the "fast flush" technique and how it can be used to estimate the fidelity of a catheter-manometer system.
12. Why is a catheter-tip transducer's lumen filled with carbon dioxide and not air?
13. Describe the Riva-Rocci method to measure blood pressure non-invasively. Why should the cuff deflation be neither too fast nor too slow?
14. Describe the Penaz-Wesseling method.
15. Why can the Penaz-Wesseling method not be used during severe vaso-constriction conditions?
16. One problem of the Penaz-Wesseling method is the unknown offset between the average pressure in the arteries and the average cuff pressure. Explain how an occasional Riva-Rocci upper arm measurement can be used to calibrate the Pefiaz-Wesseling measurements.
17. Wesseling solves the problem of measuring a finger pressure rather than the brachial pressure that clinicians want by incorporating an electronic filter that computes a brachial pressure from the finger pressure. Discuss which properties this filter should have.
18. Describe the continuous blood pressure measurement based on pulse wave velocity.
19. What is the difference between PTT (pulse transit time) and PAT (pulse arrival time)
20. Describe the challenges for calibration of a blood pressure measurement based on PTT or PAT.
21. What is the most accurate technique for measurement of cardiac output? Why is it not the gold standard?

8 Patient Monitor

A patient monitor presents meaningful information and actionable data to clinical users for diagnosis setting, guidance of therapy and monitoring of the health state during patient care. For decision making and optimizing care physicians need to have all relevant information at hand. The data quality must be of the highest level, accuracy and precision of a measurement must be guaranteed. Low error rates are extremely important. This requires a careful design of the measurements and the host device, design specifications are determined by clinical and regulatory requirements, this requires a lot of specific know how of the design team. This proprietary clinical know how is gathered in a period of many years of customer feedback and device development. It is not easy to copy this know-how but for the design of a high quality monitor this is essential. Latency (time delay between actual signal and waveforms presented at the display) must be limited (at most 0.25 seconds), the device is used in conditions where time pressure is high. The above constraints require a careful system design and a real time operating system of the host device. In summary correct decision making is possible when some data is missing, but not when incorrect data is used which is assumed to be correct. A device must present information in such a way that no confusion is possible regarding the numerical values.

The device should be easy to use even under high time pressure. This calls for a careful ergonomic design of the device and its user interface and a simple standardized method of presenting. Hence usability, configurability, user interaction and user interface are key features. On the other hand, flexibility is desirable, this conflicts with the previous requirements. Preferably maintenance should be minimal, a device should have internal calibration and error checks. Safety for the patient and users must be guaranteed for single fault conditions. In summary, reliability, usability and trusted data are the keys words. Proven and mature technologies are used in monitor design. The latest consumer and computer technology is often not robust or reliable enough for the clinical use cases.

A patient monitor is a small part of a hospital system and must be integrated in the hospital workflow and information technology system. In the ICU environment bed side data is often transmitted to central monitor stations where data from multiple patients can be monitored by specialized staff members. Devices are integrated in the hospital network and information system (HIS). The device must fit in the clinical workflow, it should be designed to minimize workload. Requirements of the monitor functionality differ per use case. There is a complete family of monitor devices each optimized for a certain use case. Devices are available for monitoring in the intensive care, during surgery, during transport, during spot checks and during emergency care outside the hospital. New concepts may be needed for low-acuity environments where patients are mobile and where the patient to nurse ratio is large.

In this chapter the technology, architecture, challenges and use in different clinical environments are described. Monitoring requirements can vary appreciably and depend on the patient, intervention and location. The requirements, measurements and standards for monitoring are described in the next section. Thereafter the architecture and block diagrams of a state-of-the-art monitor is discussed.

8.1 Monitoring Requirements

The number and type of measurements that are required for a certain use case depends on several factors. An important guideline for clinical care and thus also for patient monitoring is the principle of *do-not-harm*. The risk of a specific measurement for the patient needs to be balanced against the benefit. For instance, invasive blood pressure measurements enable continuous and accurate measurements but increase the risk for the patient in terms of complications like infections and

blood loss. It depends on the patient, the intervention and location what type of monitoring and what type of measurements are needed.

8.1.1 Patient

The patient disease severity, comorbidities and age are important factors. For instance, a patient in a shock state receives intensive care, needs mechanical ventilation, the hemodynamic status can change very rapidly and continuous monitoring of major parameters like cardiac output, blood pressure, ECG and SpO₂ is mandatory. This requires invasive measurements. On the other hand, a patient in the ward who will be discharged from the hospital in the next days may only receive a 3 time per day a spot-check monitoring where temperature, blood pressure and pulse rate are measured. A non-invasive oscillometric blood pressure suffices.

8.1.2 Intervention

The severity of the intervention is an important factor. Whenever the respiration or circulation of the patient can change quickly a higher level of monitoring is required. Anesthesia and blood loss can change the hemodynamic status of a surgical patient in very short time and when unnoticed this can have dire consequences for the patient. In this case criteria could be is anesthesia general or a local, is it major invasive surgery or a small minimally invasive intervention, the amount of blood loss that is expected, is external ventilation needed.

8.1.3 Location

Are single or multiple patients to be monitored in a room, is monitoring needed during transport, is monitoring done in an ambulance, is monitoring done during imaging in a MRI or CT scanner, are the patients mobile, can the walk around, are patients monitored at home.

8.1.4 Standards of Monitoring (ASA)

The American and British societies for Anesthesia (USA: ASA) have defined monitoring standards during surgery. During all anesthetics, the patient's oxygenation, ventilation, circulation and temperature shall be continually evaluated.

The following are minimum requirements for monitoring during anesthesia:

- Pulse oximeter (pulse rate, SpO₂)
- NIBP (automatic, every 5 minutes systolic, diastolic, mean blood pressure)
- Continuous ECG (one to three lead ECG, heart rate)
- Inspired and expired oxygen, carbon dioxide, nitrous oxide and volatile anesthetic agent if used (capnography, oxygen and anesthesia gases)
- Airway pressure and minute volume
- Peripheral nerve stimulator if neuromuscular blocking drugs used
- Temperature for any procedure > 30 min duration

Some patients will require additional monitoring, for example intravascular pressures, cardiac output or biochemical or hematological variables depending on patient and surgical factors. The use of additional monitoring is decided by the anesthetist. Use of depth of anesthesia monitors, for example processed EEG monitoring, is recommended when patients are anaesthetized with total intravenous techniques and neuromuscular blocking drugs, to reduce the risk of accidental awareness during general anesthesia.

8.2 Vital signs

A very important output of a patient monitor is the numerical value of so called vital sign parameters. Vital signs are a group of important parameters that give an indication for the status of the body vital functions. The four main vital signs are:

- Heart rate, it is measured from the ECG. Pulse rate is measured by palpation or by pulse oximetry or blood pressure. Note the difference between heart rate (indicates that the heart is electrically active) and pulse rate (measured volume changes in arteries, i.e. heart both electrically and mechanically active). There are conditions where the heart is electrically active but not mechanically. Therefore, pulse rate is the more clinically relevant parameter. The normal range is between 60 and 100 beats per minute.
- Respiration rate, the number of spontaneous breaths per minute. For an adult the normal range is between 10 and 20 breaths per minute.
- Blood pressure, measured at the systolic and diastolic level and the difference. A normal range for an adult is: diastolic blood pressure is 60-90 mmHg and 90 to 140 mmHg for systolic blood pressure.
- Temperature, i.e. the core temperature. The normal range is between 35 °C and 38.5 °C.

Other parameters which are sometimes used in the context of vital signs are pain, oxygen saturation, End-tidal CO₂ pressure and blood glucose level.

8.3 Patient Monitor Device

A high-level block diagram of a modular multi-parameter patient monitor is shown in Figure 8-1. At the patient side sensing elements are connected via shielded cables to the connectors at the patient monitor. These sensors, cables and connectors are very important for correct functioning and safety. The physiologic signals produce tiny electrical signals and electrical interference and cable motion can easily corrupt the signals. The patient interface, cabling and connectors are discussed in more detail in one of the following sections. The sensing elements are connected to the inputs of the individual measurements. The power supplies and data channels of the individual measurements are electrically isolated from both the electrical ground and mains supply according to strict regulatory standards (leakage current, ESD, KV's high voltage isolation). The measurements are also isolated from each other to prevent hazardous situations in case of malfunction of one of the measurements. The output of the measurements is a processed clean waveform and main vital sign parameters are extracted already in the measurement module. The measurement modules are connected to a proprietary data and command bus for low latency and guaranteed communication with the host processor on the main board. This data bus must be extremely reliable and flexible, i.e. it should remain functional when new measurements are added, function in harsh environments, and data transmission must be guaranteed at all time. The digital waveform output and parameters are sent to the main board and preferably do not need further processing in the host processor. Analog and digital processing and determination of vital sign parameters is often done in the measurement module. This is a good design for a modular system, measurement modules can be re-used in a wide variety of products. The data-bus control circuitry and measurement design guarantee the timing and latency of data transmission. In this way a tight control of time synchronization between the different measurements is obtained. Time synchronization of signals is very important, waveforms of different measurements need to be synchronous. This is not a trivial problem; it is a challenge and it is of vital importance that waveforms are synchronous within a few milliseconds. For example, each measurement has its own clock with its own accuracy and drift, clock synchronization is a complex problem. It is important to mention that the hardware and software of the data bus must be real

time and under the control of the host processor of the monitor to guarantee that the device meets the stringent requirements in reliability and latency. A real time operating system is mandatory.

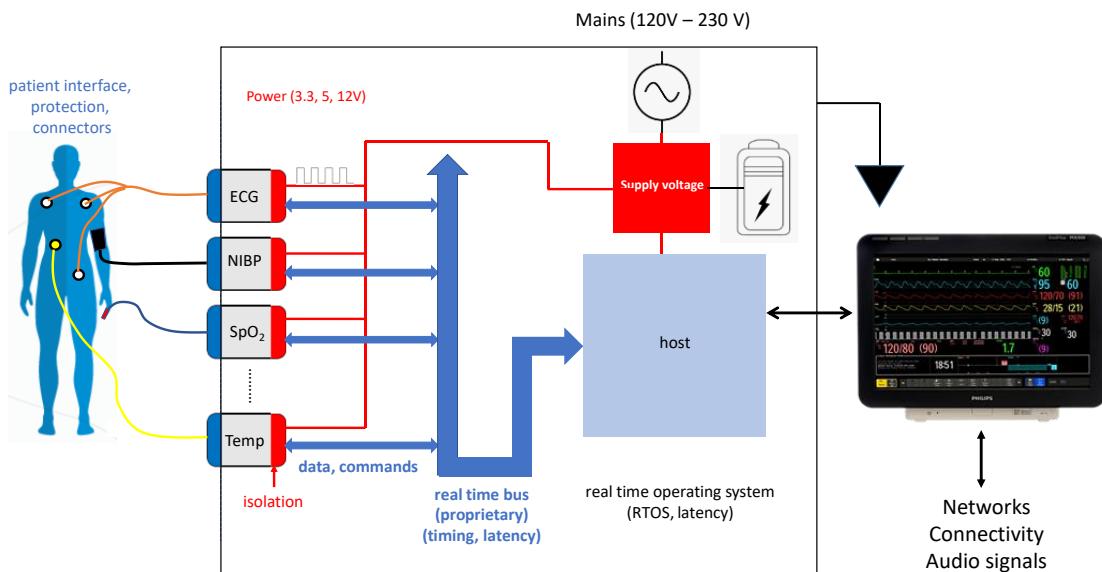


Figure 8-1 High level diagram of a patient monitor.

The data from the different measurements is in some cases further processed in the host processor. High level processing which requires more computing power is typically done in the more powerful host processor. An example is the algorithm that classifies the rhythm of an ECG signal which is very complex and requires a lot of computing power. The host processor is the system master and is responsible for alarm generation. The operating system of the host processor should be real-time to guarantee performance and latency requirements. The host processor further controls the output to the monitor display, user input, data output to external wired (LAN, Ethernet) and wireless networks (Wi-Fi, Zigbee, Bluetooth, WMTS medical wireless bands, proprietary devices).



Figure 8-2 Philips IntelliVue MX800 patient monitor with measurement rack and measurement modules.

An example of a high-end MX800 Philips IntelliVue monitor is shown in Figure 8-2. The device consists of a front-end and back-end part. The front-end is in the so-called rack module. The front-

end is the part is configurable by the user. It is the part where the measurement modules are placed, and where connections are made to the internal data busses and electrical power lines. In the left part it contains a so-called measurement server, a module that integrates five standard measurements from Philips in a single casing. Next to this device is a module for airway pressure, air flow and capnography. On the right part non-Philips measurements and anesthesia gas measurement modules are placed. In this way high-end measurements from other companies can be integrated in a controlled and standard manner to the monitor system. An example is a continuous cardiac output measurement, the PiCCO device. This front-end design makes the monitor flexible to use, guarantees backwards compatibility and is future proof. This is very important for the user as monitor devices can be used for periods longer than 20 years. Very important is the backwards compatibility, old modules should function flawless in a new system and this should hold also for using new modules in old systems.

The back-end of the monitor is the part where high-level processing, alarm generation, system control, user interaction, communication with external networks and visualization of measurements is controlled (see Figure 8-3). The back-end part is the master of the monitor and is responsible for the real time functioning. An LCD touch screen is used for display and user input. The input for patient ID and patient info should be mentioned. The Philips device is patient centric, which means that the patient monitor is linked to a specific person and not to a specific bed. The measured data belongs to a specific patient and data sent to the hospital infrastructure is coupled to the person's electronic patient file. This is very important considering that patients can be transported to other locations in the hospital and are decoupled from the monitor. When the monitor is reused again it is not obvious which patient is connected to the monitor. There is an interface to the monitor for rapid input of the patient ID. The main blocks of the monitor system are discussed in the following sections.

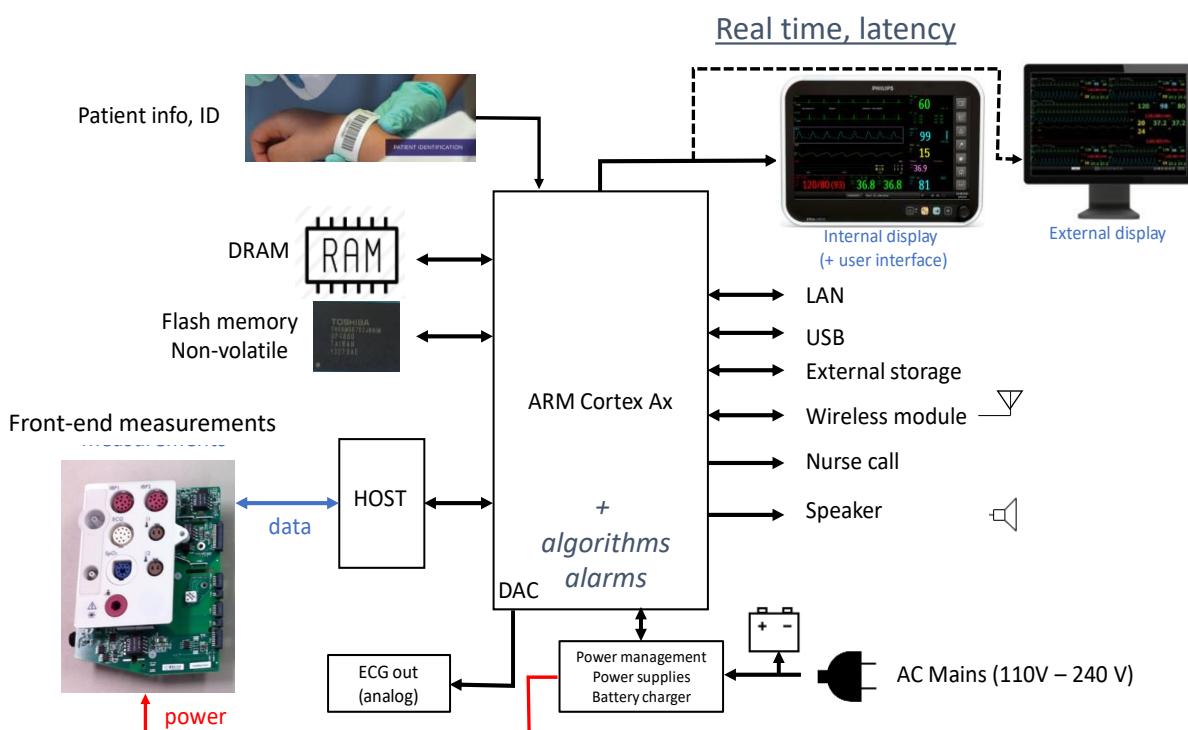


Figure 8-3 High-level block diagram of the back-end part of a patient monitor.

8.4 Patient Interface, Cables and Connectors

Signals reach the monitor via the sensing elements, the cabling and the connectors. Any additional external noise, artifact or other disturbance picked up in this part of the system will degrade the overall performance. The quality of the sensors, cabling and connectors is very important for the ruggedness and overall quality of the monitor system. Furthermore mechanical strength should be high, a rough treatment is common.



Figure 8-4 Patient side sensors, cables and connectors for an ECG and SpO₂ measurement for a Philips monitor system

As an example, consider the ECG and pulse oximetry measurements (see Figure 8-4). The ECG signal magnitude is around 1 mV and total allowed input-referred peak-to-peak noise should be smaller than 20 μ V. Coupling from mains powerlines and interference from other electro-magnetic origin can easily corrupt the ECG signal. High quality shielding and ground connections are of major importance to reduce unwanted EMC noise. Furthermore, the cables and connectors must withstand voltage pulses of thousands of volts from defibrillation devices without being damaged. Finally, leakage and creepage currents between connector pins must be smaller than 10 μ A at high kV voltages on the pins. To meet these requirements, the connectors need to have a relatively large distance between the pins, and this increases size. For pulse oximetry similar arguments hold. The dynamic range of the ppg signals is extremely large and in some cases these signals are very small. Shielding for stray light is also very important. Furthermore, the optical sensor must exert a certain contact force on the measurement site to improve signal quality. The cable is electrically shielded for the same reasons as for the ECG measurement.

The connectors for each measurement have a unique pin layout, color and shape. This prevents insertion of a sensor cable in the wrong connector and damaging the connector pins. The demands on the mechanical strength of cables and connectors are high, for example the cable should not be damaged when somebody trips over the cable or applies excessive force to push a connector into the wrong connector. This increases thickness and stiffness of the cable. The use of these cables has several drawbacks. The main one is cable clutter, the spaghetti like intertwining of cables from different measurements and ECG leads, this is one of the biggest annoyances of clinical users. At present there is no good solution.

8.5 Clinical Measurements

A clinical measurement is more than the basic measurement hardware. In this course a clinical measurement is defined as a system that has a clean output signal with derived vital sign parameters that are meaningful and actionable, the output can be used directly in a monitor (see Figure 8-5). The output of the measurement is a processed signal, the main vital sign parameters (e.g. pulse rate,

SpO₂) are also derived on the measurement board and can be used in the monitor. In the monitor more complex and higher-level processing is possible.

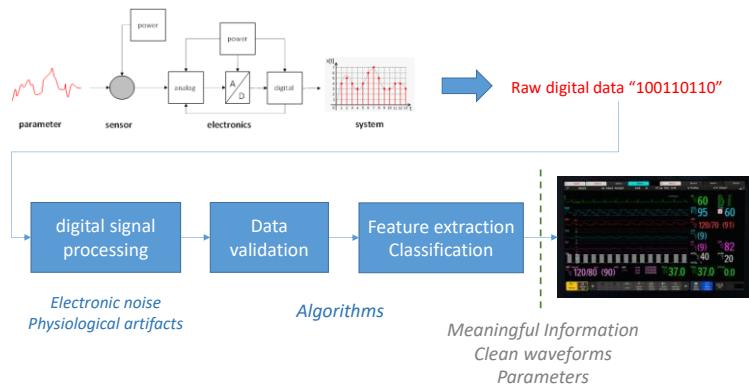


Figure 8-5 Block diagram of a clinical measurement. The measurement contains all parts up to the green dashed line.

A block diagram of a measurement PCB board is shown in Figure 8-6. On the printed circuit board (PCB) protection elements are placed before the sensitive and vulnerable electronics. The input data is filtered, and analog preprocessing and filtering is done before data are converted to the digital domain. The signal to noise ratio of the final signal is determined for a large part by the first filters and analog processing stages, the demands on the components and analog design are very high. SNR is in the range between 80dB and 100dB depending on the measurement. Furthermore linearity requirements and dynamic range demands can be excessive. Note that signals, frequencies, bandwidth and number of bits are comparable to the requirements for high-end digital audio. The digital data is fed into a microcontroller (MCU). The MCU controls the measurement and does digital signal processing, removes artifacts and determines vital sign parameters, finally data is transferred to a digital bus. The processing power of modern ARM M4 core MCU's is often much larger than is needed for subsequent digital signal processing. Complex digital filtering and digital signal processing is used for noise and artifacts removal, data validation and extraction of features and vital sign parameters. The MCU has standard digital bus interfaces like SPI or I2C. Since byte rate is not excessive simple serial interfaces are used. Drift in local clock time must be controlled and synchronization of the measurement with a central clock is mandatory.

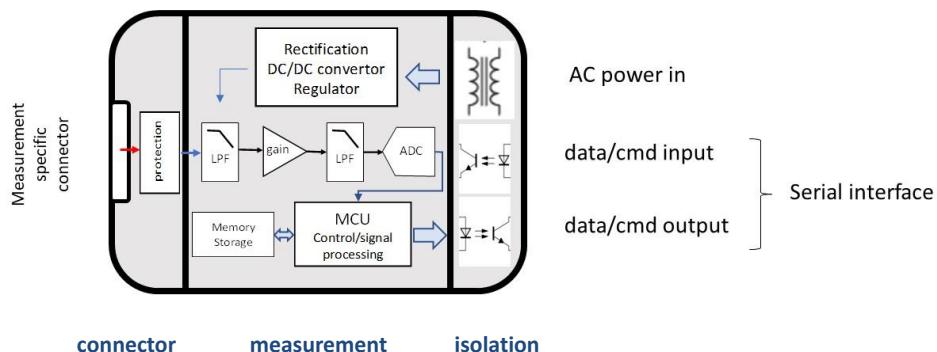


Figure 8-6 High-level block diagram of a clinical vital sign measurement.

The measurement needs to be galvanic isolated from the mains supply (kV's), the monitor ground connection and central board electronics. Isolation transformers are used to transfer electrical power to the main board. Isolation requirements are in the range of 5kV or higher, this requires space and

volume and limits miniaturization. The powering and multiple voltage generation occupies a large part of the measurement board. The AC signal from the isolation transformer is rectified and voltage converters generate multiple voltages for the analog and digital components. Clean and stable DC voltages in the range of 1V to 12V are required. Note that although the power efficiency of these components is high (> 80%), the serial combination of converters can reduce efficiency levels and increase power overall consumption. This is of special interest when power consumption needs to be reduced, e.g. portable devices. Data and commands are sent to and from the measurement board. These data connections are also electrically isolated from the mains supply and main board of the monitor. Optical isolation is used most frequently. The requirements on voltage isolation are similar to those of the power section. Note that the protection, filters, voltage and power conversion and isolation cover a large part of the PCB area and device cost.

8.5.1 Data Processing and Parameter Extraction

Although the problem which data should be acquired and analyzed is application-dependent, in modern integrated monitoring systems the basic signal processing procedures will often be comparable. Algorithms perform the acquisition and processing of a standard set of measurements and extract features: maxima, minima, periods, slopes, rates etc. Due to the increase in the capacity and performance of modern integrated circuits the analysis and processing of large amounts of data is at present less an issue. The quality of the algorithms is a dominant factor and often influences the selection of a measurement of a certain brand by the customer. Two problems have a general character. The first is that the acquisition rate of many of the measurements is so high that no human would be able to handle the 'raw' data; some sort of data preprocessing is required so that only more meaningful data will be offered at a much lower rate. The second problem is that the quality of the data is to be suspected. Due to a variety of causes, the acquired data may not reflect the quantity that they are supposed to represent. A process of *data validation* is required to establish the authenticity of the acquired data. Signal processing must isolate the clinically significant features of the signal; data validation may be based on some additional features as well.

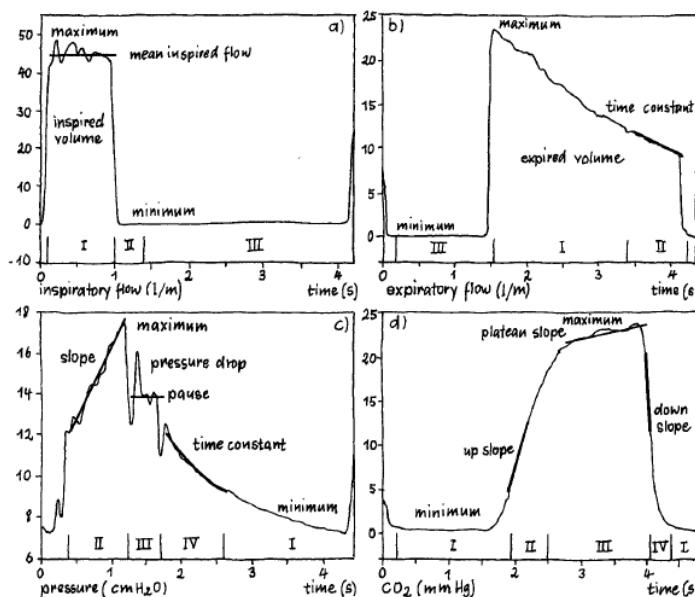


Figure 8-7 Measured inspiratory (a) and expiratory (b) flow, respiratory pressure (c) and CO₂ concentration (d) signals frequently deviate from the "ideal" signals generated by a model. Dog experiment.

Figure 8-7 shows an example of some actual "raw" respiratory signals measured in a dog. The curves are far less "clean" than the idealized curves that have been presented thus far. Yet, the idealized curves

are guides that tell us which features should be extracted. The inspiratory flow signal must yield the maximum, minimum and mean flows, the inspired volume and the inspiration period. The expiratory flow signal must similarly yield the maximum, minimum and mean flows, the expired volume and the expiration period; in addition, an exponential is fitted to the decreasing part of the curve in order to obtain the "time constant", the product of expiratory resistance and lung-thorax compliance. The pressure signal must produce the maximum, minimum and mean pressures, as well as the stepwise initial pressure increase, the inspiratory slope, the inspiratory pause pressure, and the time constant of the expiratory part of the curve. The latter parameters allow some of the properties of the respiratory circuit to be estimated. From the CO₂ signal, minimum and maximum concentrations as well as several slopes must be determined. Data processing algorithms often need to perform simple operations only: discover discrete points (maximum, minimum, starts and ends of slopes) in the curve or in its low pass filtered version, fit linear or exponential functions to signal sections (slopes, time constants), or determine averages (means). Some algorithms need to perform much more complex operations, such as the pattern recognition that is required to discover abnormal QRS-complexes in an ECG signal or the estimation of SpO₂ in a pulse oximetry measurement.

Figure 8-8 shows a schematic of a classical single signal monitor. A transducer acquires the signal. Because no transducer is perfect and exactly reproducible, some type of correction may be required, e.g. for non-linearity or fluctuations in characteristics from device to device. In an invasive arterial pressure measurement, for instance, a calibration sequence may be needed in order to generate an "ideal" signal that is independent of an individual transducer's offset and gain. The offset calibration can be combined with the "null" calibration, which is required anyway to eliminate the hydrostatic pressure component after a manometer has been put into position. Part of the correction may be filtering to eliminate noise and unwanted signal frequencies.

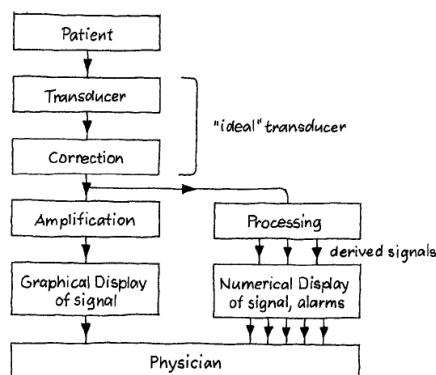


Figure 8-8 Data processing steps. A signal is acquired by a transducer whose characteristics may need correction. After "dumb" but faithful amplification, the signal is presented as a graph. After possibly very complex signal processing, the signal's features are presented as a number or a set of numbers.

After correction, the signal goes two ways. First, it is visually presented onto a monitor display with a latency less than 0,25 seconds with the actual signal, because the signal's shape contains information about measurement- and patient-related problems. Second, the signal is processed and its derived features (maximum, minimum, period, rate etc.) are presented as alphanumeric. The values of the derived features are also compared with alarm limits; when these are exceeded, a visual and/or auditory alarm is generated.

Figure 8-9 provides a conceptual model for the processing of data in a modern, computer-based patient monitoring system. The data is acquired by the basic measurement devices, as depicted in the top part of Figure 8-7. The data validation process determines whether the data are valid or if it is an artifact. In the latter

case a warning could be generated. If the data contain considerable redundant information, e.g. in case of a waveform, the feature extraction algorithms extract all meaningful features from each period of the signal, as in the rightmost part of Figure 8-7. If one or more of the extracted features are abnormal, the feature classification process attempts to determine the type of abnormality. The trend analysis process classifies and analyzes the dynamics of the data. The history extraction process builds a compact history of the feature over e.g. the last few hours. The problems that must be solved determine the information that must be available in the database. This information in turn determines which signals must be acquired and which features must be extracted from these signals. The characteristics of the signals in turn determine how the features must be extracted.

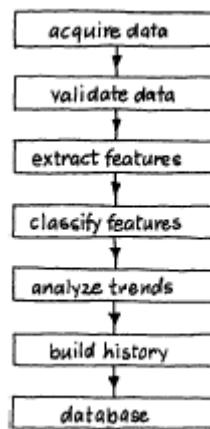


Figure 8-9 Data processing steps in a modern patient monitoring data acquisition system.

8.5.2 Data Types

The monitor has multiple functions. As mentioned in previous lectures the device functions as the sensor in the patient care process. It is important that the monitored data is stored. Waveform data can be stored in the monitor for a 24h or 48h period. It is deleted automatically after this period. During patient care and monitoring several types of data are relevant, part of this data is not gathered by the monitor (e.g. fluid administration, medicine administration, context information, imaging results, lab data, etc.). The combined data is very important for the clinician, a review of the data serves as important input for further diagnosis and treatment plans. This type of data is stored in the hospital information systems. Furthermore, it is essential to store the data for future analysis and record keeping and research purposes. This combined data set may be of importance for future treatment, legal processes and clinical studies.

A main function of the monitor device is that it provides continuous data to the clinicians and warns them for imminent instabilities in the patient condition. A warning or alarming function is one of the most important functions of the monitor. The following data types are encountered.

8.5.2.1 Demographic Data

The monitor data belongs to a patient. It is important that patient identity and demographic data (age, sex, weight, length, ...) and data from the specific monitor, other measurement data and imaging results are entered in the unique patient data file of the hospital information system. The patient ID is entered in the patient monitor before or directly after the patient is connected to the monitor. This is often done electronically by scanning a bar code that is attached to the wrist of the patient. This identification process is not trivial, errors are common and missing data is a major problem. This is also relevant for large scale data analysis and data centric studies, data that cannot be trusted must be labeled.

8.5.2.2 *External Data*

During surgery or stay in the ICU frequently medicines or anesthetic agents are administered. Blood samples extracted from the patient are submitted to the laboratory for subsequent analysis.

Administration of drugs and lab test data must also be recorded, and a time stamp and description of the substance and dose must be recorded. These intermittent or external events are very important for outcome and must be annotated and combined with the monitor data. Examples are:

- Injections of drugs (drug type, dosage)
- Infusions (drug or fluid type, infusion flow rate, bolus dose)
- Fluid loss (blood, urine)
- (Changes in) settings of ventilators, gases, flows
- Interventions (emergency care, stop blood loss, change of invasive lines, ..)
- Blood sampling, blood gas analysis, measurement of substances.
- Observations of the clinicians
- Context information
- Imaging results (ultra-sound, x-ray, MRI, CT, ...)
- Addition of new (continuous) measurements (e.g. adding a Swan Ganz catheter for cardiac output)

Some external devices have electronic outputs that can be coupled into modules for external data entry of the monitor. For instance, most mechanical ventilators have internal monitoring of flows and gases and this data can be transferred to the monitor either in analog or digital format. Waveform traces from the ventilator can be observed on the monitor display. However, in many cases data cannot be transferred to the monitor. Until recently these processes were manually annotated, which is prone to errors. Recently annotation using smart devices is introduced, note that most of the annotation is still manual. Due to time pressure this annotation is often delayed and data may not be reliable. It is possible to add time stamps to the monitor to mark the timing of an event. As noted before under time pressure this may have no priority.

This external data and information is very important during and after the patient care process. It is essential during patient care to optimize the treatment. For subsequent analysis and research this information is crucial, when it is missing or in error the post hoc analysis may be meaningless or even misleading.

Presently hospital information systems are upgraded, and all patient related data and patient information should become integrated in a single system that can be consulted by the medical staff. This process is very complex due to the wide variety of information, data and systems that must be integrated. Presently the standardization of these processes is not optimal, and a large effort is required to improve the situation. Furthermore, there is an intense competition between the different companies and organizations involved, often having different (proprietary) standards for data storage. The integration of these multiple data sources is one of the main challenges and time-consuming activities for the hospital IT organization.

8.6 Monitor Display

The display of the monitor is the link between the transducer, measurement results (sensor output) and clinician. The display has a touch screen functionality and serves as the user interface where functionality and monitor settings can be set or adapted. The ergonomic design of the display, its user interface, contents and lay out are important. The display should be a simple as possible while still be able to show the fine details when needed. The user interface must be intuitive and change of settings and parameters must be possible in a short time. The design of the user interface

influences the speed in which information can be acquired by the clinician. An example of the screen of a Philips monitor is shown in Figure 8-10.

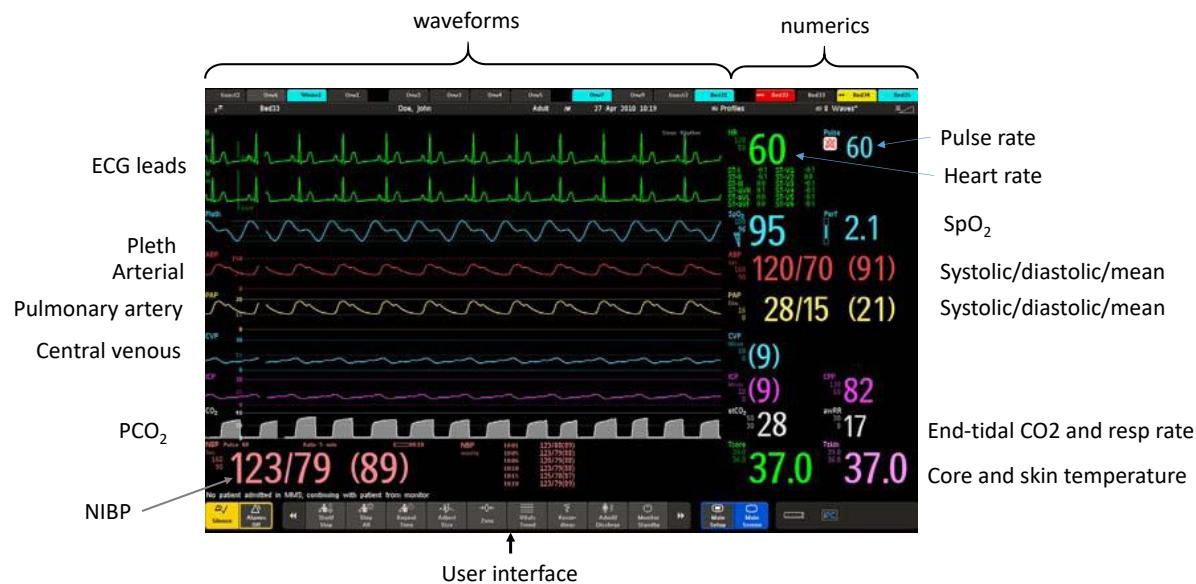


Figure 8-10 Display of a Philips IntelliVue MX800 monitor.

Waveforms of the parameters are shown in the left and central part of the screen. They contain “analog” information that is not present in the numerical data. This information is very valuable and can be used by the experienced clinician for diagnosis and treatment. Furthermore, it indicates the quality of the measurement and gives trust in the value of the extracted numerical (vital sign) parameters. Error messages/warnings from the measurements are displayed, for instance high contact resistance of the ECG electrodes or poor sensor signal quality. This electronic warning of the signal quality is presently still inferior to the analysis of the waveform by a trained clinician. The time axis of the waveform area is around 12 seconds and waveforms are continuously refreshed. The time delay between the actual patient signal and the displayed waveform must be less than 0.25 sec. This is needed for treatment during emergency situations. Each parameter and related waveform have the same unique color. Numeric (vital sign) data is plotted on the right part of the display. A moving average value is plotted rather than the actual value. There is often a time variation in the beat to beat values, this continuous variable data distracts the user. The main buttons of the user interface are displayed at the bottom of the display. Via a menu the structure settings can be changed quickly.

Optionally other information can be shown on the display (see Figure 8-11). Trend plots of parameters are of great interest for the analysis of the patient condition and decision making. Trends and changes in parameters are often more important for the clinician than the actual numerical data. In recent years clinical decision support data and plots can be seen on the display. The example shown in Figure 8-11 is a spider plot of values of the ST elevation in a multi-lead ECG measurement. It helps clinician in a faster analysis and shortens reaction time. In the right an integration of an ultrasound image in the monitor screen is shown. The fusion of imaging and monitor data is a long-standing wish of the clinician. Another option is to show plots of one variable versus another variable. An example is a pressure-volume loop during positive pressure ventilation.

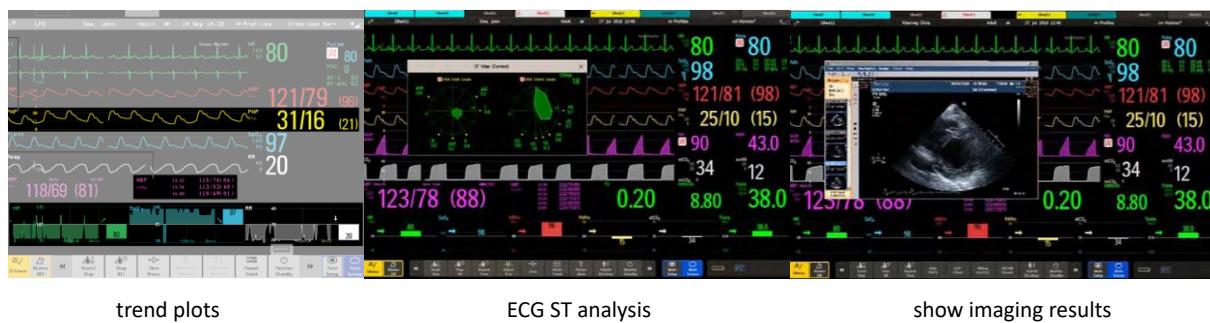


Figure 8-11 Monitor display with trend plots of parameters, clinical decision support plots of an analysis for ST elevation and integration of imaging data.

8.7 Alarming

Alarms are intended to warn the clinician for imminent and potentially harmful conditions in the health state of the patient. There are different type of alarms. They are illustrated in Figure 8-12 in a time series of the heart rate. There are typical ranges of vital sign parameters that are considered normal. For instance, for heart rate the normal range is between 60 and 100 beats per minute. Whenever the heart rate is lower or higher the monitor gives an alarm. The monitor will produce acoustic and optical signals that are difficult to ignore. The default settings for these parameters are factory set but the clinician is responsible for the actual values. Note that the context is very important, a normal heart rate can be much higher than 100 bpm when the patient is moving. Note also that vital sign parameters during disease states can differ a lot from textbook values and can also differ from baseline values of the specific patient. What is normal and what is acceptable should be the decision of an experienced clinician and is time and patient dependent.

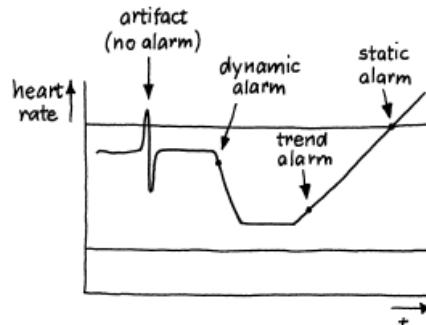


Figure 8-12 Diagram that shows a time series of the heart rate and alarm limits. The different type of alarms are indicated in the figure.

The simplest alarm is the static alarm, whenever a parameter exceeds the set limits an alarm is generated. A static alarm has limited information, the clinician must take further action or ignore the alarm whenever the clinician decides it is not important or is due to an artifact (i.e. moving cables, moving patient). The static type of alarm is most common in monitors but has a very high rate of false alarms. More than 90% of the alarms are false positives. Another type of alarm is the dynamic alarm, the parameter changes rapidly but may still be in the acceptable range. The rate of change is the trigger parameter. Finally, a long-lasting trend could be indicative of a forthcoming problem (trend alarm). Note that most alarms are caused by artifacts, device malfunction or to tight alarm limits, the alarms are not clinically relevant and distract the clinician.

A more precise classification of alarms is listed below.

- Detection of immediately threatening conditions. The origin of all alarm types.

- *Asystole, VF, VT, Hypoxia, Apnea, ...*
- *Standard in all monitoring systems*
- Detection of (life) threatening device malfunction
 - *Disconnection from patient*
 - *Disconnection from supplies (energy, gas, water)*
 - *Internal malfunction*
 - *Standard in all therapeutic devices*
- Detection of incipient danger
 - *Before it leads to a life-threatening state*
 - *Trends, Arrhythmia, ...*
- Detection of imminent device malfunction
 - *Before it leads to a life-threatening failure*
 - *Low battery power, mechanical wear and tear, ...*
- Detection of pathophysiological states
 - *Diagnostic alarms*
 - *Interpretation of data (e.g. Hypovolemia)*
 - *Extension to decision support systems*

The first type falls in the category of static alarms, it is however of the highest level of danger, immediate action is needed, every second counts. When the heart stops beating, or the patient stops with breathing immediate action is required. The second type is the equivalent form for a therapeutic device. When a ventilation machine is disconnected from a patient it is a life-threatening condition, the lung ventilation stops. The third category is the dynamic or trend alarm. It is a form of an intelligent alarm. The alarm level is lower than that of the first two alarms, immediate action may not be needed. The fourth alarm type is a warning for imminent device malfunction, low battery power is a typical cause, disconnection of sensors to the patient occurs frequently. Poor ECG electrode contact resistance is another example of such an alarm. The last type is a more intelligent and smart alarm. Complex features and trends are extracted from the monitor signals, it can also include a possible interpretation of the data. An example could be the detection of sepsis by combining data from several measurements. A next step could be clinical decision support which supplies the clinician with patient specific information and options for further decision making and treatment.

The most frequent device and physiological alarms are shown in Figure 8-13. The most dominant technical alarms are for the SpO₂ and ECG measurements. Often trivial causes such as sensor detachment, poor sensor or electrode contact, moving cables and motion artifacts are causing problems. Furthermore, a low signal to noise ratio may be a problem. The post processing algorithm may have problems with feature extraction due to poor signal quality or poor performance of the algorithm for that specific waveform. Automatic detection of cardiac rhythms is notoriously difficult. For the physiologic alarms the most frequent types are high respiration rate and low SpO₂. Many cardiovascular, respiratory diseases and infections cause acidosis (lower pH) that leads to fast and shallow breathing. The high rate or respiration rate is not specific for a certain disease. When bio-impedance is used for measurement of respiration rate motion and movement of cables can lead to errors. This can often not be detected by the monitor system. For this reason, there is a low trust in data determined from bio-impedance. Note that the frequency of low blood pressure related alarms is not high. Blood pressure is a late indicator of cardiovascular instability.

Technical Alarms:

Rank	Event Text	Alarm Count	Percentage of Total Alarms
1	SpO2 Non-Pulsat.	2931	35.09
2	Cannot Analyze ST	1215	14.55
3	Resp Leads Off	738	8.84
4	SpO2 Searching	708	8.48
5	Ecg leads off	650	7.78
6	SpO2 No Sensor	609	7.29
7	Cannot Analyze Ecg	460	5.51
8	IntelliVue in Standby	333	3.99
9	No Current IntelliVue Data	236	2.83
10	RA Lead Off	111	1.33
11	LL Lead Off	109	1.31
12	Ecg Noisy Signal	87	1.04
13	SpO2 Sensor off	69	.83
14	V Lead off	42	.5
15	LA Lead Off	29	.35
16	RL Lead Off	25	.3
Total		8352	

Physiological Alarms:

Rank	Event Text	Alarm Count	Percentage of Total Alarms
1	RR High	1819	22.71
2	SpO2 Low	948	11.84
3	RR Low	853	10.65
4	HR Low	727	9.08
5	CVPm High	434	5.42
6	ARTm Low	400	4.99
7	SpO2 DeSat	350	4.37
8	NBPm Low	348	4.34
9	HR Low	337	4.21
10	Apnea	245	3.06
11	NBPm High	204	2.55
12	PAPd High	183	2.28
13	ABps High	146	1.82
14	NBPs High	134	1.67
15	Extreme Tachy	131	1.64
16	ABps High	127	1.59
17	NBPs Low	106	1.32
18	Pulse Low	69	.86
19	Pulse High	61	.76

Figure 8-13 Typical causes for device related and physiological alarms.

The number of alarms can be very high in an ICU or cardiac care unit. Frequently over 10000 alarms are generated per day. Most of the alarms are generated by patient monitors but the large variety of therapeutic devices each with an own alarm function add to the cacophony of sounds. Studies have shown that more than 90% of the generated alarms are false positives [Imhoff]. The large number of non-relevant alarms combined with the large workload eventually leads to ignoring of alarms. This non follow-up or suppression of the alarms jeopardizes patient safety. This abundance of false alarms leads to so-called alarm fatigue. This has been recognized as the number one patient safety hazard by the US ECRI institute. “From a clinical perspective, major improvements in alarm algorithms are urgently needed” [Imhoff].

Missed alarms, suppressed alarms and non-follow up of alarms can have dire consequences for the patient. Alarm fatigue is a complex multifactorial problem that cannot be solved easily. A main factor is the present implementation of alarms on patient monitors. The devices are designed to maximize the sensitivity¹³ of the alarms. *The number of false-negative alarms should be very small.* The alarm limits are set at levels such that sensitivity is close to 100%. The consequence is that there are many false positive alarms. The specificity¹⁴ of the alarms has been sacrificed for this purpose. In summary it is more important to reduce the number of false negative alarms; this goes at the cost of a high false positive rate. This problem has been studied for many years and there are solutions that reduce the false alarm rate, however the number of true positives is also reduced. Even small reductions in the true positive rate are a problem. The legal requirements imposed by the regulatory organizations and liability of the hospital organization are other important factors. Finally, the legal liability of the responsible doctor often forces them to avoid risks and the alarm limits are set so that the chance of missing an alarm is minimal. The result is a very high number of false alarms. The clinical staff must respond to the alarms but is often overloaded. One consequence may be that after a few false alarms the next real alarm will be ignored. At present the legal issues hinder implementation of smart algorithms, the liability of supplier is the main issue.

¹³ Sensitivity is the conditional probability of correctly detecting a condition and generation of an alarm when that condition is present.

¹⁴ Specificity is the conditional probability of correctly not detecting that condition when that condition is not present.

Whenever a new alarm algorithm is introduced the supplier must prove that this does not lead to a degradation of the sensitivity. This is not an easy task, as described above a higher specificity is accompanied by a lowering of the sensitivity. A second factor for the high alarm rate is the design of the user interface. Often setting of alarm limits is not straightforward, is blocked to the user (often by purpose, alarm limits should be changed only by highly qualified staff) and requires training of the users on the specific monitor. Therefore, alarm trigger levels are mostly not optimal for the specific patient and limits are set (too) wide. Moreover, alarm management strategy is often not implemented in the hospital and time for training is scarce due to the high workload. The issue of false alarms has been studied for a long period and it has been shown that relatively simple smart alarms could reduce the alarm overload, could increase the response rate and follow up rates and improve the overall patient safety. A few examples are discussed below (see Figure 8-14). -----

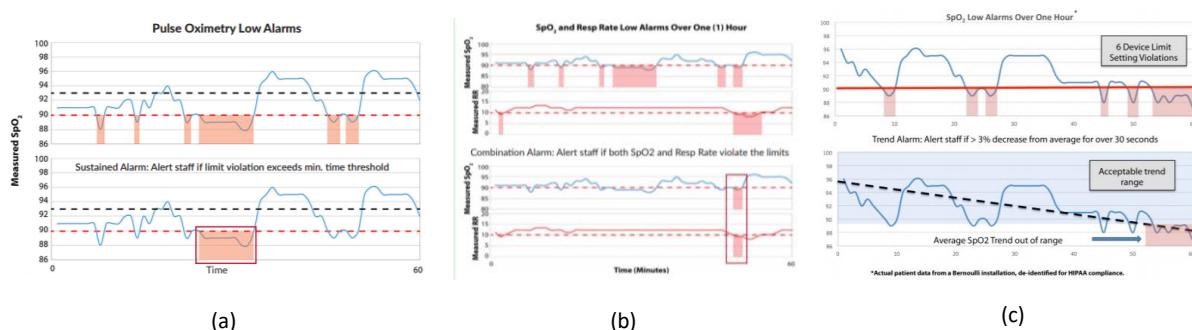


Figure 8-14 Three methods to reduce the number of alarms. (a) Specifying the minimum duration of the alarm condition. (b) Multi-parameter alarm generation. (c) Smart trend alarming.

A simple method to reduce the alarm rate would be generate an alarm only when a limit has been crossed for a specific time (see Figure 8-14 (a)). This is implemented in existing monitors, but regulatory requirements limit the duration of the time interval to 10 seconds. A longer period of 20 seconds would be much more effective. Studies have shown that a reduction of 80% in the number of alarms is possible for this longer duration. Note that for life threatening alarms such as cardiac arrhythmias a much faster alarm generation must occur. A related method is that of alarm delay. A combination of delay and alarm severity (i.e. deviation from the set limits or type of alarm) could greatly reduce the alarm rate.

A more complex method is to generate an alarm based on a combination of parameters. An example is shown in Figure 8-14 (b). Here both SpO₂ and respiration rate is monitored. Only when both parameters exceed limits an alarm is generated. Other example could be to combine related physiological parameters. When for instance the ECG measurement indicates an asystole, but invasive blood pressure shows a pulse and blood pressure is in the acceptable range it is likely that there is an undetected error in the ECG measurement. A warning signal could be generated and follow up of the alarm is less time critical.

Another approach is to generate a score that is based on the values of multiple parameters and monitoring the trend in the score. Whenever the trend indicates that the limit will be reached soon a warning could be generated. This is an early warning score system. An advantage is the clinicians can attend the patient before a life-threatening condition is reached. This system is used in low acuity environments like the ward. However, in this case a warning is generated and not an alarm as this system is designed to detect slow deterioration of the health status and not a life-threatening condition.

Another more advanced approach would be to use signal processing of time series and advanced analysis of data of a period of a certain length (e.g. 10 to 20 seconds) to remove noise and artifacts from the time series [Imhoff]. This would reduce the alarm rate because many alarms are generated by such artifacts. An even more advanced procedure would be to generate intelligent alarms that are based on algorithms and/or models of the underlying physiology. These methods are used in other fields than medicine such as nuclear power plants and airplane engines. Furthermore, for a long period machine learning techniques (random forests, Bayesian networks, neural nets, rule based expert systems ...) have been applied to study generation of false alarms. Although improvements in the reduction of false alarms have been reported most studies are retrospective, there is no breakthrough yet. Finally, a lot of research is done on diagnostic alarms that would not only generate an alarm but that could also indicate the cause of the alarm and possible options for treatment. This is part of clinical decision support which is a hot topic.

It is important to note that to gain acceptance of these advanced methods the algorithms should be transparent, and the underlying science should be comprehended by the user. Trust in the methods require that the “raw¹⁵” data must be available to the user. The waveforms on the display remain essential. It is interesting to see how machine learning techniques could contribute to this complex problem. These techniques may detect new information in the signals that is presently not used in clinical decision making. Note however that the remark on transparency to the clinical user will remain important.

8.8 Family of Patient Monitors

A hospital is a complex organization which consists of departments with a specific function and specialization. A diagram of hospital departments where patients are monitored is shown in Figure 8-15. The type of monitors used in these departments depends on the requirements and is described below. The Philips IntelliVue family is used as an example.

The departments where the highest level of continuous monitoring is needed are the operating room (OR) and the intensive care unit (ICU). These patients have serious illness and require a high level of critical care. External ventilation, fluid and medicine administration are common. Many therapeutic devices are present. The health condition can change quickly, high-quality invasive measurements of the circulation and respiration systems are needed.

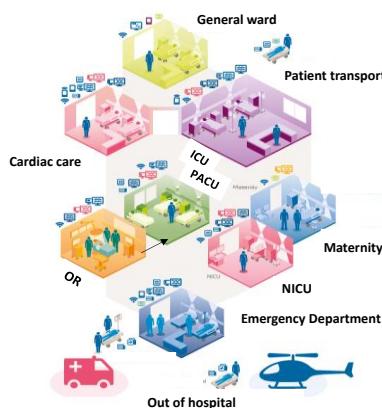


Figure 8-15 Schematic diagram of departments in a hospital where patients are monitored.

¹⁵ Raw data is not displayed on a monitor, the waveforms on the display are processed signals. The waveforms should still be visible, processing level should be as low as possible.

High-end monitors such as the MX800 and MX700 are used in these areas (see Figure 8-16). The integrated PC (iPC) option enables the display of a large amount of clinically relevant information from the hospital intranet and data from application software that is present in the hospital. Imaging results can be shown, and high-end clinical decision support applications can run on the embedded PC. Specialty measurements from external suppliers such as a continuous cardiac output or depth of anesthesia (BIS) can be inserted when they are needed. These specialty measurements can be added in the rack module. Device interfacing modules and modules for digital input of third-party devices (monitors, ventilator, anesthesia, infusion pumps) can be added. Finally, an X2 or X3 transport monitor can be added to the rack in the same way as a multi measurement module. The X3 device can be used as a stand-alone monitor and supports other use cases such as monitoring during transport and other phases of the care process (imaging, therapies, and recovery). When coupled to the high-end monitor the waveforms and numeric data are copied to the main display.



Figure 8-16 High-end patient monitors of the Philips IntelliVue monitor family with a rack module for adding of specialty measurements.

After surgery patients are brought to the recovery room (PACU, post-anesthesia care unit)). External ventilation is stopped, and a weening process is started to promote spontaneous breathing. Patients must be monitored continuously until the vital sign parameters indicate that the patients are stable and can be transported to a lower level of care. In a few percent of the cases the patient condition degrades, and a higher level of care is needed. These patients are transported to the ICU unit. The requirements of monitoring are lower and typically MX550 and MX400 monitors are used (see Figure 8-17).



Figure 8-17 Midrange MX500 IntelliVue patient monitor during use and showing the modularity and connectivity options of the device.

Some patients are transported to other departments and need continuous monitoring. The example of a patient that is transported to, or from the OR has been mentioned already. Other examples are transport from the ward to the ICU, transport of ICU patients to imaging modalities, transport of patients from the emergency department to the OR is another example. A monitor device that can travel with the patient to the new destination is a transport monitor. In the IntelliVue family a small device is available (X3 monitor). It is based on a multi-measurement server; the monitoring function and display have been added. It is shown in Figure 8-18, the predecessor X2 device connected to a MX500 monitor is shown in Figure 8-17.



Figure 8-18 Philips IntelliVue X3 transport monitor used during transport and connected to the rack module of a high end monitor.

This device offers the same measurement quality as the larger devices and reduces the workload of the nurses as sensors and cables must be connected only once. The device can run on the battery for more than 6 hours. The measurements on the X3 are preconfigured.

Most hospitals have a dedicated cardiac care unit (CCU) where patients are treated during and after a heart attack or during recovery after cardiac surgery. This unit has special requirements such as 12 lead ECG monitoring. Typically, an MX type monitor is used at the bedside. During recovery patients are encouraged to walk around and will have physiotherapy, this improves the recovery and outcome. These mobile patients need to be monitored, the risk for cardiac arrhythmias is still large. For this use case *telemetry devices* have been developed. These are small patient worn monitors that can measure three to five lead ECG, pulse oximetry and oscillometric blood pressure can be added. The devices stay with the patient even during showering. Mechanical requirements on water resistance and drop strength are high.



Figure 8-19 Philips IntelliVue MX40 telemetry device used on mobile patients and with adding cableless NIBP measurement.

Since devices are patient worn, they need wireless connectivity to the central CCU monitoring unit. This is discussed in a following section.

In the CCU and in the rest of the hospital the risk for cardiac arrest is high and defibrillator devices are present on the departments. On top of this there also monitor-defibrillator devices which are

used by the resuscitation teams in the hospital (see Figure 8-20). These devices combine the function of an advanced monitor and a high-end defibrillator. Typical monitoring functions are up to 12 lead ECG, invasive and non-invasive blood pressure, pulse oximetry and capnography. The devices can also guide CPR by measuring chest compressions and monitoring depth and frequency of the compressions and guide the user to comply with the latest recommendations of the resuscitation councils.



Figure 8-20 Philips MRx and Efficia monitor-defibrillators.

When patients are on the ward the level of monitoring is strongly reduced. Typically, three times per day a nurse visits the patient and performs a spot checks. The nurse has a short discussion with the patient and measures pulse rate, temperature and blood pressure. Recently the respiration rate has been added. These manual measurements are often not accurate or reliable and reporting is mostly on paper. Studies have shown that many adverse events happen on the ward, it is recommended that a large group of ward patients need to be monitored more closely. Quality of care and follow-up of a suspicion of imminent deterioration of the health status must be improved. To enhance measurement quality and to promote digital recording and storage of the measurement data spot check monitors are being used in some hospitals. These devices can trigger follow up actions via a system like the Guardian solution. An example of the use of a spot check monitorin the ward is shown in Figure 8-21.



Figure 8-21 Philips IntelliVue MP5SC spot check monitor in use during a spot check.

The monitor is mounted on a cart and travels with the nurse. Patient ID is checked, and measurement time and date and measured data are sent via the wireless network to the hospital information system. Simple scoring systems are used to trigger further actions. Monitoring on the ward is much more complex than previously assumed, nurses have little experience with monitoring,

there is a wide variety in health conditions, patients are often mobile, the monitoring should be unobtrusive, the rate of false alerts and warnings is high and the organization of the hospital has to support this process. Very important, the process should not affect the workflow and workload of the nurses. Recent studies have shown that many adverse events occur in the ward and monitoring on the ward and other low acuity locations will be discussed in a separate lecture.

Many out-of-hospital patients enter the hospital via the emergency department. This can range from people who come on own initiative to patients with highest trauma or cardiovascular disease levels via the ambulance or even the trauma helicopter. Outside the hospital monitor-defibrillator devices are the standard of monitoring. Inside the hospital the EMS department used midrange MX type monitors and transport monitors.

8.9 Central Monitoring Stations

In many ICU departments the monitor data of all patient monitors is redirected to a central monitoring or viewing station. A specially trained healthcare provider evaluates the situations and if necessary, notifies the nursing staff for evaluation and treatment. An example is shown in Figure 8-22. This reduces the alarm burden for the nurses at the bedside and improves the quality of care.

Besides the real time monitor data, it is possible to show data from the hospital information system and higher-level applications (clinical decision support). Similar central viewing stations are available for the cardiac care unit and recently also for the hospital ward (Philips Guardian System). Warnings, alarms and other data can be viewed on smart devices such as tablets and smart phones. Recently services have become available to outsource the central viewing monitoring station and provide expert advice to the ICU of smaller hospitals. Remote monitoring, video surveillance and expert support can be delivered from remote locations. With central monitoring real time warranted connectivity must be available. This is discussed in the next section.



Figure 8-22 Central monitoring stations and solutions.

8.10 Hospital Networking and Connectivity Solutions

Patient monitoring devices send data via the hospital wired and wireless network systems to the central station and hospital information system. In Figure 8-23 a diagram of the network connections between an IntelliVue monitor and the central station and the hospital information system is shown. For the wired network connections local area network (LAN, Ethernet) connections are used.

Monitor data is time critical and low latency must be guaranteed. Data and alarms sent to the central station must have a guaranteed quality of service. A proprietary LAN network optimized for use in a hospital guarantees that data and alarms arrive at the central station with minimal delay. Data and other information are sent to HIS according to the HL7 standard (Health Level 7). It is apparent that the monitor device is a part of a larger eco system that comprises many other devices, data storage and software services and standards.

A telemetry device is a kind of patient worn monitor and has therefore no wired LAN connection to the hospital wired network. Telemetry devices use a wireless link to the hospital network. This is shown in more detail in Figure 8-24. In Figure 8-24 (a) the different options for use of wireless connected monitors are shown. The wireless transceivers in the monitors communicate with wireless access points which control the communication between the various devices that are in the range of the specific wireless access point and are connected to the central fixed LAN network of the hospital. The first example is a periodic monitor such as a spot check monitor. Short bursts of numeric data are sent to the central monitoring station and HIS via the fixed LAN network. This is not time critical and latency is less an issue.

The second example is the use of a portable monitor which sends both numeric and waveform time series data and alarms via the access point and fixed network to a central monitoring system. An example is a transport monitor. Wireless transmission is not as reliable as a wired (proprietary) LAN network. Since alarms are transmitted the information is time critical and quality-of-service is a main issue. In this case a clinician is nearby (transport) and local alarming on the monitor suffices in case the network connection is not present.

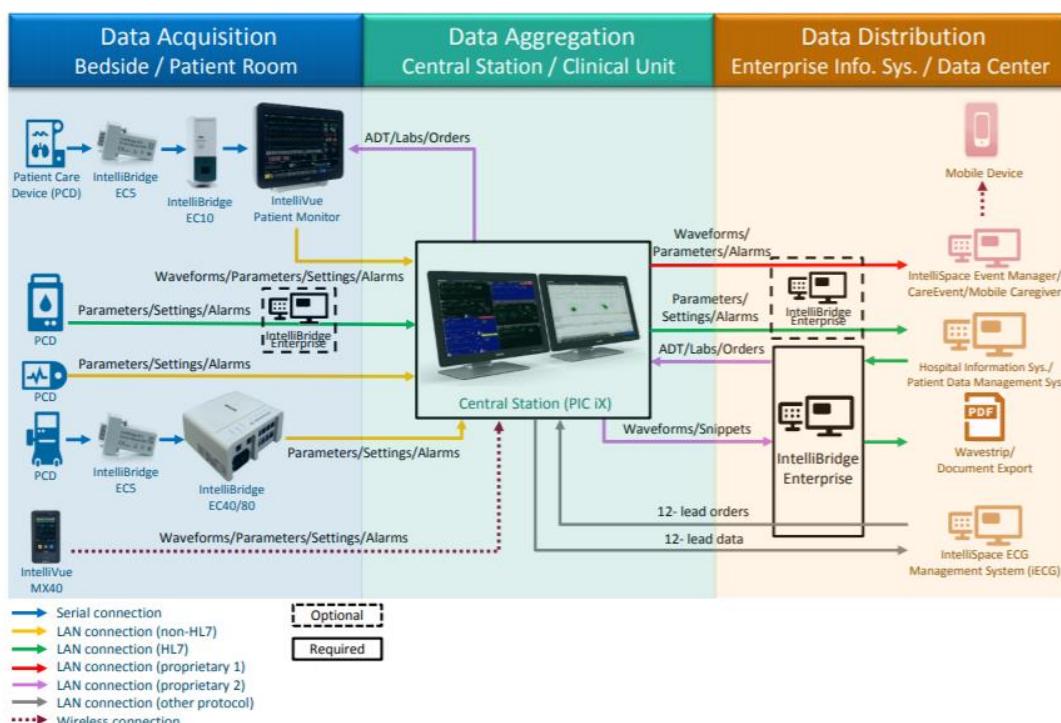


Figure 8-23 Diagram of connectivity between monitors, patient care devices and central station and HIS with wired and wireless networks in a hospital.

The third use case is the most challenging one. The device is used as an ambulatory monitor. The patient can be alone in a room, walk through the hospital, be in the restaurant or in the bathroom or the toilet. In this mobile patient case, the device alarm signal must reach the central station. Patient monitoring devices require transmission of real time signals with low latency. This cannot be guaranteed for all cases when wireless transmission of data is the only way to connect to a central monitoring system. In short, WiFi and Bluetooth wireless data transmission is at present not 100% reliable and can therefore at this moment (2022) not be used in cases where latency is important.

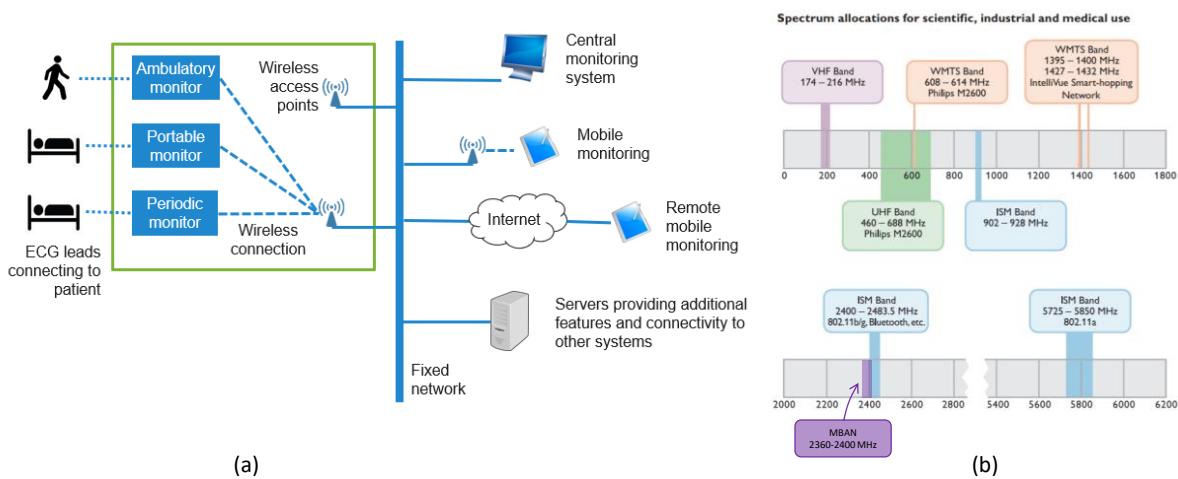


Figure 8-24 (a) Diagram of wireless connectivity options and use cases in the hospital. (b) Most used spectrum allocations and transmission standards.

Optionally monitor data and alarms can be transmitted via the LAN and Wi-Fi network to clinicians with smart devices or to remote locations via the internet. Data can also be sent to servers and data storage centers for storage and higher-level processing. Clearly the quality and reliability of the wireless system is very important for these use cases.

In patient monitoring applications, many devices are sending continuous data into the infrastructure at the same time, all the time. When public networks like WLAN (Wi-Fi) are used for transmission of medical information the situation is even more complex. These networks are often congested by data from other users and visitors of the hospital. For a wireless link quality of service cannot be guaranteed, the quality of service depends on many factors.

The first factor is the wireless standard and frequency band that is used. Is it an open frequency band or is it restricted to medical use? The options are illustrated in Figure 8-24 (b). Old telemetry devices used a part of the VHF and UHF bands that were not used by digital television. This is an efficient solution when there are few competing devices. In case of interference a complete channel can be blocked, or latency can be unacceptably large.

A small part of the wireless spectrum in the USA is reserved for WMTS (Wireless Medical Telemetry Service). It is a protected spectrum for transmission of life-critical data (physiological parameters and other patient-related information) in healthcare facilities. Equipment operating in the WMTS bands (1395-1400 MHz, and 1427-1432 MHz) is operating under primary status, and therefore, protected from interference by other devices. This is the major advantage. Note that a WMTS network is a proprietary network system that must be installed specifically for the telemetry system. This high cost(> 100k€'s) comes on top of the cost of the telemetry devices. Furthermore, the data rates of the WMTS system are optimized for patient monitoring devices developed in the 1990's, higher data rates (e.g. needed for images, video) are not possible. Data cannot be sent directly to smart devices, these are compatible with WLAN and Bluetooth only. In addition, only few suppliers of WMTS transceiver devices are available.

The ISM (Industry Science Medical) bands are large, unlicensed areas of spectrum for a growing variety of devices that can be used to transmit everything from ECG waveforms to multimedia news streams. The ISM bands are also used for WLAN (Wi-Fi, IEEE 802.11), Bluetooth and ZigBee devices. Wi-Fi and Bluetooth are integrated in mobile devices such as laptops, tablets and smart phones.

Therefore, these bands are often overcrowded, and the capacity and latency cannot be guaranteed. WLAN networks are available in most hospitals and are part of the existing infrastructure.

Recently a Medical Body Area Network (MBAN) standard has been proposed. This standard developed for a body network of small wearable sensors and protected transmission to the hospital system. The allotted frequency band is between 2360 and 2400 MHz. This band is (under limitations) available in the USA and Europe. This standard is still in the initial phase, components are not available, and interoperability is an issue.

The idea of using one 802.11 wireless infrastructure for patient monitoring and non-medical devices is appealing, especially for hospitals that already have a network in place. Using the ISM band for this type of application monitors must compete with the many other smart devices that use the same spectrum. Effective throughput could then be degraded, and latencies throughout the system would increase and worse cannot be guaranteed. Since physiologic data has “real-time” requirements, excessive latency cannot be tolerated. The perceptible effect of latency for end-users is gaps in the waveform display at the central station, gaps in wave review and discontinuities in arrhythmia monitoring.

Nevertheless the use of Wi-Fi connectivity for patient monitoring applications is growing, the technology is now mature and stable, high-quality enterprise systems are available and the 5GHz band offers new capacity and more channels. Some suppliers of telemetry devices are focusing on Wi-Fi only or are adding Wi-Fi solutions for telemetry. This is often done in cooperation with a network supplier to optimize the wireless infrastructure in the hospital. This is needed as in many hospitals the WLAN networks do not function properly and maintenance is not optimal.

Good results can be obtained when network capacity and quality of the network system is managed as a life-critical infrastructure. This is the second important factor for the quality-of-service. This depends on the quality of the installed hardware, the network management and maintenance and the capacity and quality of the hospital IT department. This brings additional costs to the hospital, and experienced IT experts that need to manage the system are often not available. The trend for wireless connectivity of telemetry devices is at present towards Wi-Fi. The technology momentum and industry support are very large and this technology is more compatible with smart devices in use in the hospital and would also allow real time video transmission or imaging results.

8.11 Summary

The market for patient monitors is a mature and very conservative market. There is less emphasis on new vital sign measurements. A better utilization of data is needed. There is a demand for algorithms that extract more information, provide better classification and reduce artifacts, and support the clinician. For lower acuity monitoring there is a demand for less obtrusive measurements, body worn measurements and lower power especially for the patient worn devices. Safety, ease of use, reliability, interoperability remains difficult and challenging. A monitor is a very complex device with very specific requirements for use-ability, reliability, latency and safety for the patient and user. A monitor is part of an eco-system of devices, IT systems and software applications and services. The link to the Hospital Information System is increasingly important for administration, quality, legacy and research studies. Monitors have a long lifetime (10-20 years) and new systems should preferably be backward compatible. New use cases are emerging, both for conventional ICU and OR applications down to low acuity to monitoring basically everywhere. A major issue remains alarming, the rate of false alarms is far too high. Fortunately, data analysis, artificial intelligence and remote services offer opportunities to resolve this longstanding issue provided that legal issues are resolved. Trends in patient monitoring and solutions will be discussed later.

8.12 References

[Webster] John G. Webster, Medical Instrumentation, Application and Design, John Wiley and Sons, fourth edition, ISBN 9780471676003.

[Blom] J.A. Blom, Monitoring of Respiration and Circulation, CRC Press LLC, ISBN 9780849320835

[Imhoff] M. Imhoff et al., Smart alarms from medical devices in the OR and ICU, Best Practice & Research Clinical Anaesthesiology 23 (2009) 39–50.

8.13 Questions

1. Discuss the conflict in instrument design requirements that medical measurement device manufacturers face.
2. Discuss the conflict in information presentation requirements by medical measurement devices.
3. Describe the process through which a clinician reaches a diagnosis.
4. Give an example how measurement methods' perceived risk and utility may influence the choice of a method.
5. Briefly describe the type of measurements that are most important in a) a coronary care unit, b) an intensive care unit, c) an operating room.
6. Discuss the notions of normality and stability for physiological measurements.
7. What do "data validation" and "signal validation" mean?
8. Discuss Figure 8-7. What shape would the model-based "ideal curves" of in- and expiratory flow, respiratory pressure and exhaled carbon dioxide have?
9. Discuss the data processing steps of a typical medical patient monitor (Figure 8-9, Figure 8-8). What is the function of each step?
10. Discuss the types of data that an integrated patient monitoring system must be able to acquire and store.
11. Compare the approximate data storage requirements for systems that store 20 physiological measurements sampled at 250 Hz. Assume the following storage strategies: a) the samples themselves, b) 80 derived variables reduced to a rate of 1 Hz; c) 80 derived variables stored once every 30 seconds.
12. What are some practical guidelines for the design of a visual display layout?
13. What is meant with the terms "static alarm", "dynamic alarm", "trend alarm", and "intelligent alarm"?
14. What is a clinical measurement, how does it differ from a physical measurement?
15. What are the main causes for device and physiology related alarm generation.
16. Describe the architecture of a modular multi-parameter patient monitor.
17. Describe the issues with wireless connectivity in a hospital environment.
18. Smart phones and tablets have a much larger computing power and a much nicer display and have Wi-Fi and Bluetooth connectivity. Discuss why such devices cannot fulfill the main functions of a patient monitoring system.
19. What are the factors that limit integration of multiple measurements on a single integrated circuit?
20. Discuss the main issues and challenges of existing patient monitor devices.

9 Therapeutic Devices

Therapeutic devices support or replace critical physiological functions. An example is a patient ventilator that replaces the spontaneous respiration process of lung ventilation when the patient is not able to breath. Other examples are infusion pumps for administration of medicines, anesthesia agents and pain treatment, heart-lung machines that replace the function of both the respiratory and circulatory system during high level cardiac or pulmonary surgery, pacemakers that support or replace the function of the SA node, defibrillators that revert life-threatening cardiac arrhythmias in perfusing rhythms. Therapeutic devices often include a monitoring function for both physiologic and device related parameters and are used together with patient monitor devices to observe the effect of therapy and adjust therapy when needed. The basics of mechanical ventilation devices in use in hospitals for respiratory therapies is described in this reader. The other devices will be discussed briefly, in the appendix defibrillators and infusion devices are discussed in more detail.

9.1 Mechanical Ventilation

When patients undergo major surgery in the operating room, general anesthesia must be applied. Anesthesia suppresses the central nervous system, causes unconsciousness and most sensory functions are stopped. Respiratory muscles are relaxed and the patient is not capable of spontaneous breathing (see Figure 9-1). The lung function of surgical patients function is in general normal, however this may change during surgery. A mechanical ventilation device is connected to the patient to maintain lung ventilation.

A second important use case of mechanical ventilation is the patient in the ICU unit. Mechanical ventilation replaces or supports lung ventilation in patients in critically ill patients. These patients are not able to breath independently anymore. Common causes are heart failure, shock states, systemic infections and respiratory failure (sepsis, pneumonia, COPD, ARDS, COVID-19, ...).



Figure 9-1 Anesthesia, mechanical ventilation and monitoring during surgery.



Figure 9-2 ICU patient in prone position with mechanical ventilation support. Prone position improves the ventilation-perfusion ratio of the diseased lungs.

The goal of mechanical ventilation is to maintain gas exchange with blood that sustains life and buys time for further treatment and recovery. In these cases a ventilation machine is needed to replace or support the spontaneous breathing for the patient groups mentioned above. Often the lungs are (partly) affected by inflammation or severe infection, are filled with fluids, certain areas in the lung are collapsed and are not ventilated anymore (atelectasis). A ventilated COVID-19 patient in the ICU is shown in Figure 9-2. The patient is in prone position to improve ventilation and oxygenation of blood. In general the lungs of ICU patients are very vulnerable and can be damaged easily when too large airway pressure or too large tidal volume is used. Mechanical ventilation of ARDS or COVID patients is complex and potentially harmful for the patient. It requires training, skills and experience

to optimize and personalize patient ventilation during the stay in the ICU. The basics of mechanical lung ventilation are discussed first.

9.1.1 Airflow to the Lung

During the inspiration phase the pressure at the airway opening must be larger than the lung pressure. This can be realized in two manners as is shown in Figure 9-3. The first is the oldest type of ventilation.

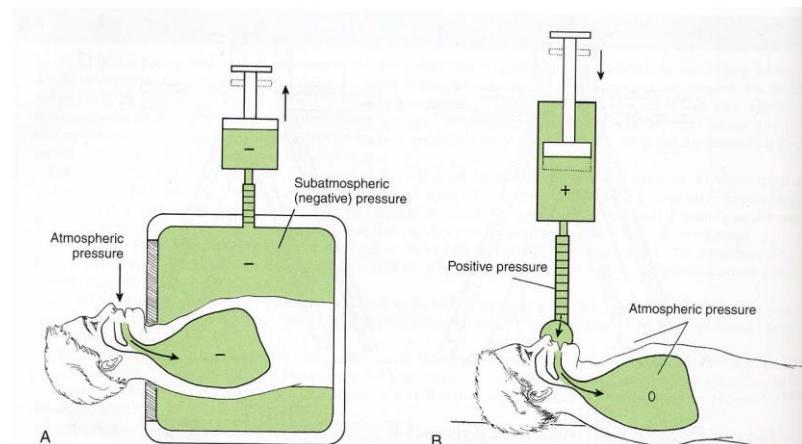


Figure 9-3 Schematic drawings of negative body pressure ventilation (A) and positive airway pressure ventilation machines (B).

In spontaneous inspiration the muscles of the patient exert a negative pressure on the lung and the lung pressure is sub-atmospheric, the airway opening (AWO) is at a higher atmospheric pressure. Air flows from the AWO into the lungs until the lung pressure equals AWO pressure. In a negative pressure ventilation (NPV) machine this process is mimicked. The external thoracic body surface is exposed to sub-atmospheric pressure (i.e. negative pressure) during the inspiration phase. The mechanism of lung ventilation is the same as for the normal spontaneous breathing, lung pressure is slightly negative. Expiration of air is passive. In this system lung pressures are low and lung injury is avoided.

As can be seen in Figure 9-3 (A) the patient body surface is enclosed by a rigid tank, the neck is sealed with a rubber gasket. The head of the patient is in the room with atmospheric pressure. The pressure in the tank is varied by a piston device or a pump. Patients can talk and eat normally. The NPV device has many drawbacks. Note that the whole body surface is exposed to sub-atmospheric pressure, this has consequences for the blood distribution in the body. Blood will pool in the large veins and blood flow to the right heart is reduced. External massage of the legs and abdomen is required to resolve this problem. Furthermore washing and cleaning of patients and removal of urine and feces is labor intensive. Finally such systems are not suited for patients with respiratory failure where controlled over-pressure on the lung is needed.

The NPV method was developed in 1929 and used on a large scale in the period of 1950 to 1980 for polio patients who required long duration lung ventilation. As noted above it has many drawbacks. From the 1980's negative lung ventilation machines are replaced by positive pressure airway ventilation (PPV) machines.

The basic concept of Positive Pressure Ventilation (PPV) is shown in Figure 9-3 (B). The patient body surface is at atmospheric pressure. At the end of an expiration the lung pressure is near atmospheric pressure. When the pressure in the airway opening (AWO) is raised above this end-expiratory pressure air flows into the lungs. Note that high-pressure air in the mouth can flow both in the

esophagus and trachea. The airflow in the esophagus (D) is harmful and can be lethal when it is not detected. Therefore an endotracheal tube with a lumen diameter around 7mm-10 mm is inserted in the trachea(C) (see Figure 9-4). A balloon is present near the end and can be inflated via port B. The balloon blocks reverse air flow into the esophagus (D). When an endo-tracheal tube is used it is called invasive ventilation. When the tube is inserted in the esophagus the lungs are not ventilated and air pressure in the stomach is large. This is a life threatening condition. Proper insertion of an endotracheal tube requires training and skill. The use of capnography during tube insertion reduces the risk of misplacement. A drawback of the endotracheal tube is that it adds considerable to the airway resistance due to turbulent airflow in the tube at normal flow velocities. The tube flow resistance can be (much) larger than that of the airways of a healthy adult person. Furthermore the tube can induce injuries in the mouth and throat during prolonged use. Finally the patient must be sedated as the procedure is very uncomfortable to the patient.

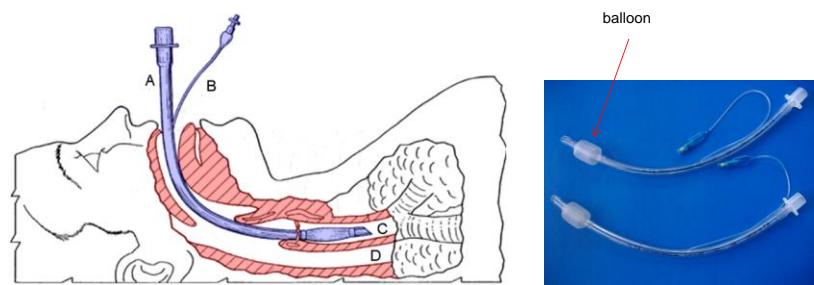


Figure 9-4 Drawing of an endotracheal tube (A) inserted into the trachea (C). A balloon can be inflated via port (B). When the balloon is inflated air cannot flow back into the esophagus.

After intubation the ET tube is connected to a Y-piece and breathing circuit, the connection with the ventilator (see Figure 9-5). The breathing circuit consists of an inspiratory tube (fresh air) and an expiratory tube (expired air). Both are connected to the Y-piece. Distal of the Y-piece and close to the airway opening proximal sensing devices for pressure, flow and capnography can be placed. The sensors are connected to the respiratory monitor that is included in most modern ventilator devices and parameters, pressure, flow and volume waves are displayed continuously. *The proximal sensing position is preferred because flow resistances of the breathing circuit and endotracheal tubes are large compared to the airway resistance of an average adult and measurement errors when pressure is measured in the ventilator would be too large.* Furthermore the compliance of the tubes cannot be neglected, part of the ventilator flow and tidal volume is used for inflation of the flexible tubes.

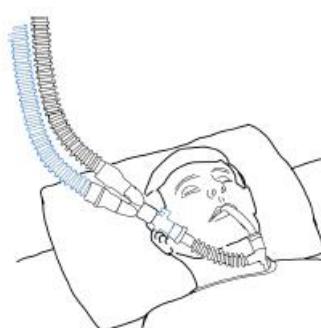


Figure 9-5 Schematic drawing of patient connected to a breathing set of a PPV mechanical ventilation machine.

Valves located in the ventilator control the flow into (inspiratory valve) and out of the patient (expiratory valve). There are different modes of ventilation.

A schematic diagram of the pressure, flow and volume waveform of a flow-controlled ventilator is shown in Figure 9-6. For simplicity we assume that the inspiratory flow is a block wave and the airway resistance and compliance of the lung of the patient can be approximated by simple constant lumped elements. Important parameters are indicated in the figure.

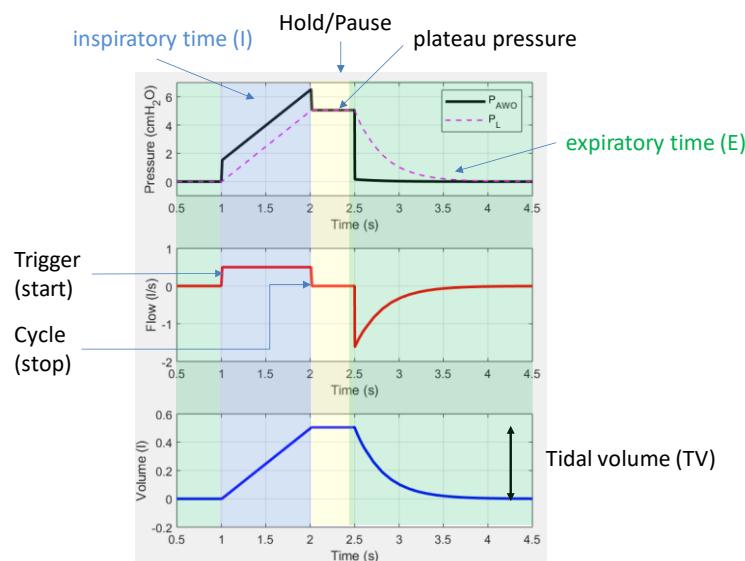


Figure 9-6 Pressure, flow and volume curves during one flow controlled PPV cycle. Important parameters are indicated.

The first parameter is the start of a ventilation and is called trigger. The inspiratory time (I) is the time interval when air flows into the patient lungs. The time when the inspiratory valve closes and air flow is stopped is the cycling time (cycle off). The time integral of the inspiratory flow is the tidal volume. During inspiration the airway pressures rises due to the pressure drop over the airway resistance and an increase in lung pressure due to the increase in lung volume. As an option at the end of the inspiratory time both the inspiratory and expiratory valves can remain closed for a short time, the airway is occluded, this is the hold or pause time. No air flows into the patient and resistive pressure drop over the lung airways vanishes and the airway opening pressure is equal to the lung pressure. During the hold time the pressure in the airway opening is called plateau pressure. This is the maximum lung pressure during ventilation. It is related to lung injury due to excessive pressure (barotrauma) of the lung, maximum inspiratory pressures need to be limited. Furthermore the hold period can be used to estimate the lung compliance and airway resistance. After the hold time the expiratory valve opens and air flows from the lungs to the expiratory port of the ventilator and the expiratory phase (E) starts. This is a passive process. The flow waveform has an exponential shape. The time interval between opening of expiratory valve and opening of the inspiratory valve is the expiratory time (E). The expiratory time should be longer than three or four times the RC time of combined patient-ventilator circuit to prevent incomplete expiration and over ventilation. An important parameter is the ratio between inspiration and expiration time, the so called I/E ratio. A typical value is 1:3 to 1:4. The main ventilator techniques and simple models of the patient-ventilator system are discussed next.

9.2 Ventilator Modes of Ventilation

First so-called mandatory modes of ventilation are discussed. There is no patient effort and the ventilator controls the airflow into the patient. There are two basic positive pressure ventilation

(PPV) ventilator modes. In the first mode the inspiratory pressure is the controlled parameter and airflow is the dependent variable. In the second type the flow and/or tidal volume are controlled and the airway pressure is the dependent parameter. One speaks of pressure controlled (PC) or flow or volume controlled ventilation (VC). A simple ventilator model and a one lung model is used to illustrate the two methods.

9.2.1 Pressure Controlled Ventilation

In Figure 9-7 a schematic diagram of a pressure controlled system is shown. The ventilator model includes two ideal pressure sources, one for the inspiration pressure pulse and one to deliver a minimum positive-end-expiratory-pressure (PEEP), in this example PEEP is set equal to atmospheric pressure. PEEP is an important parameter that is discussed later.

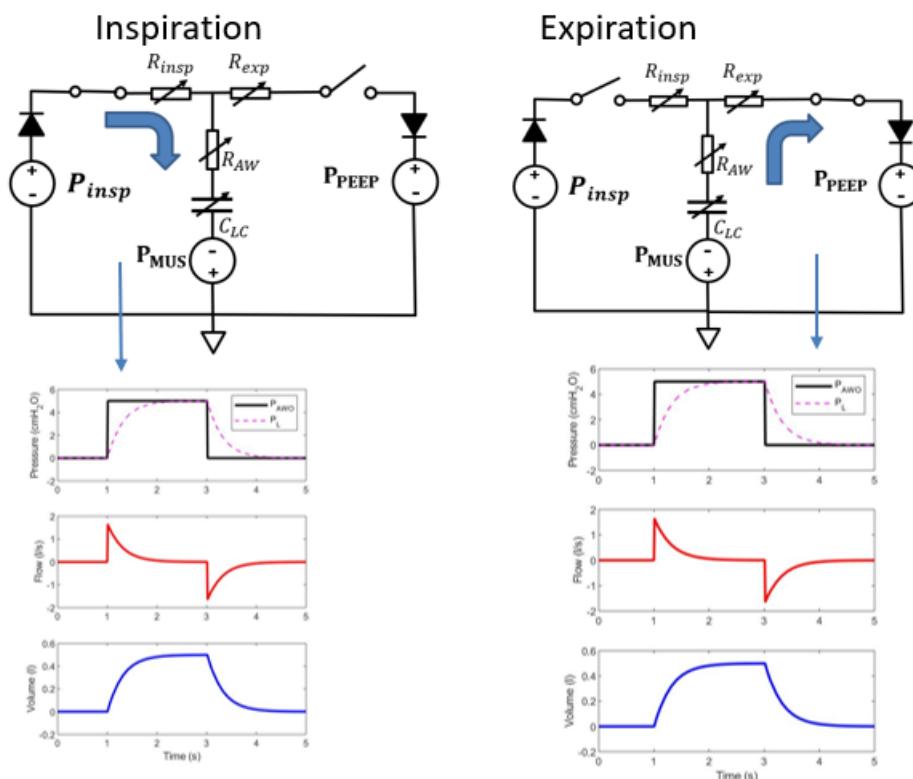


Figure 9-7 A ventilator generates a inspiratory pressure P at the inspiratory valve position for an inspiration time T_i . R is the airway resistance and C is the total lung-thorax compliance. The dashed line is the pressure drop over the lung.

There are two one-way valves (modeled as diodes) to guarantee unidirectional flow from the inspiratory port to the patient and from the patient to the expiratory port. The ventilator is modelled as a simple pressure source and two valves. The tube resistances from the inspiratory and expiratory tubes in the breathing circuit are included, they are large and cannot be neglected. The patient model consists a resistor to model the airway resistance and one compliance which is the combined lung-chest wall compliance. A muscle pressure source is added to model the negative pressure that is exerted on the lung and chest wall by the respiratory muscles and lung chest-wall elasticity. Both the airway resistance and lung-chest wall compliance are in practice non-linear. *In this example a linearized model is used, resistances and compliances are constants and muscle pressure is zero, i.e. muscles are relaxed, there is no patient inspiration effort.* Pressure and flow are measured in the ventilator.

The ventilator generates a controlled pressure P wave at the inspiratory port, for simplicity we assume a simple block wave. In Figure 9-7 both the inspiratory and expiratory ventilator and lung pressures and the airflow and volume waves are shown. The solution for time dependent pressure and flow are the same as for voltage and current in a low pass RC filter i.e. a low-pass first-order system (see appendix 13.2.3). The lung pressure is the node pressure at the lung-chest wall compliance. The time constant of the system is equal to: $\tau = (R_{insp} + R_{AW}) \cdot C_{LC}$. The resistance is the sum of the breathing set and airway resistances. The flow has a sharp peak just after the inspiratory pressure is raised and decays exponentially until the lung pressure is equal to the inspiratory pressure. The lung pressure increases exponentially with time towards the limiting inspiratory pressure. The lung volume follows the plot of lung pressure. After three time constants 95% of the maximum tidal volume is reached. The tidal volume depends on inspiratory pressure, inspiratory time and parameters of tubing and the patient respiratory system. The tidal volume is dependent on the inspiratory pressure and time and is determined by the integral of the flow over time. Hence the tidal volume depends on patient and device parameters. It is not controlled in this example. At the end of inspiration the inspiratory valve is closed, the expiratory valve is opened and air flows from the lung towards the ventilator. The exhalation is passive, the elastic recoil force from the patient drives air out of the lungs. The expiratory flow has the reversed sign, peaks directly after the expiratory valve is opened and varies exponentially with time. Lung pressure and inspiratory volume also decrease in an exponential manner. The peak flow and exponential rise and fall of the flow are characteristic of a mechanically ventilated passive patient during pressure control. The time constant τ is an important parameter for the clinician, it provides information about lung parameters. In the following section the flow/volume control mechanical ventilation is described.

9.2.2 Flow/Volume Controlled Mechanical ventilation

A schematic diagram of the patient-ventilator model and the three main phases of flow-volume controlled mechanical ventilation are shown in Figure 9-8.

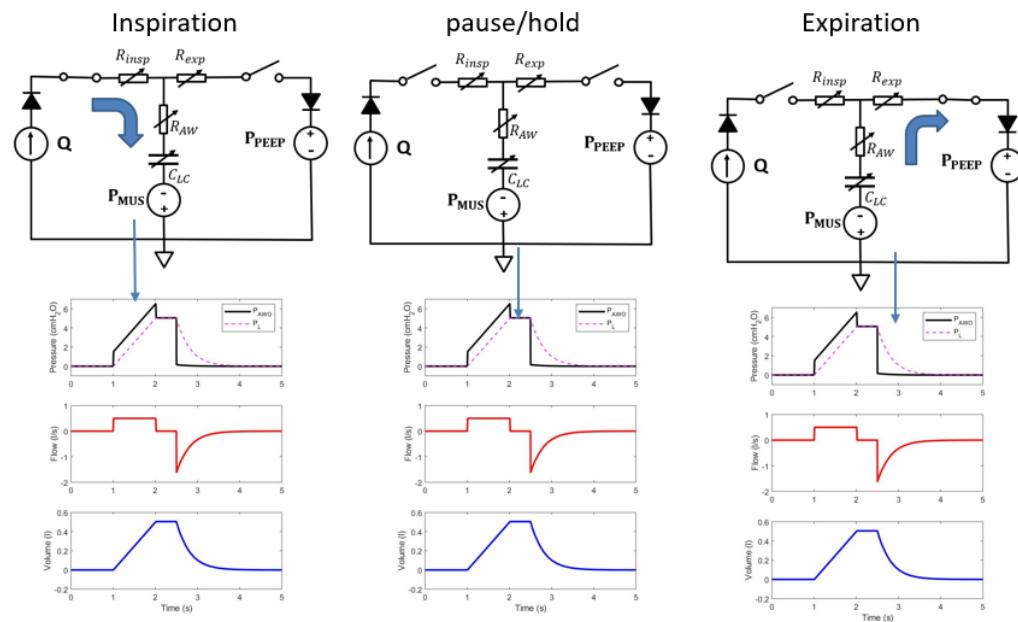


Figure 9-8 Schematic diagram of the mechanical ventilator, its valve positions and the airway pressure, flow and volume waves is shown for the inspiration, hold/pause and expiration conditions. Pressure is measured in the ventilator.

During inspiration phase the ventilator generates a constant flow Q . Then the pressure at the airway opening at time t during inspiration is given by:

$$P(t) = Q \cdot R_{tot} + \frac{Q \cdot t}{C_{LC}} = Q \cdot R_{tot} + \frac{TV}{C_{LC}}$$

9.2-1

The inspiratory pressure, flow and volume waves are shown in Figure 9-8. They are obtained from equation 9.2-1 which is based on a solution of the simple series RC network. Directly after opening the inspiratory valve, the measured inspiratory pressure increases with a step, this increase is caused by the Ohmic pressure drop over the total resistance (i.e. the sum of tube and patient airway resistance). This is the resistance component of the pressure wave. Thereafter the pressure increases linearly with time, this is the increase in pressure over the lung-chest wall compliance, it is caused by the constant flow Q into the compliance C_{LC} . This is the elastic pressure component (dashed line). The tidal volume TV is equal to the product of inspiratory flow and the inspiration time. The increase in lung volume with time is linear as the lung-chest wall compliance is constant.

During the hold time both the inspiratory and expiratory valves are closed, the pressure drops suddenly, this is the pressure drop over the tubing and patient airways. The pressure is constant during the hold/pause period for this simple system. There is no flow and the lung and airway pressure are equal, this is the so-called plateau pressure. Note that the tidal volume TV is controlled precisely as both the flow Q and inspiratory time are known. The tidal volume is controlled and does not depend on the tube resistances and physiologic patient parameters.

After the hold time the expiratory valve is opened and expiration is due to the passive recoil force of the elastic lung-chest wall. This phase is identical to that of the pressure controlled ventilation. The pressure, flow and volume waves are identical to that of the pressure controlled case. The exponential decay of both pressure and flow is characteristic of the passive expiration process.

When there is a hold/pause time during flow controlled ventilation the resistance and lung compliance parameters of the lung can be estimated. The peak inspiratory pressure (PIP) is the maximum airway pressure, the plateau pressure is the pressure during the hold/pause period and is equal to the lung pressure at the end of the inspiratory period. For a given tidal volume and flow rate the PIP and plateau pressures cannot be controlled. The pressure depends on resistances and patient physiologic parameters. Controlling the maximum pressure is important for reducing lung injury. This is discussed later.

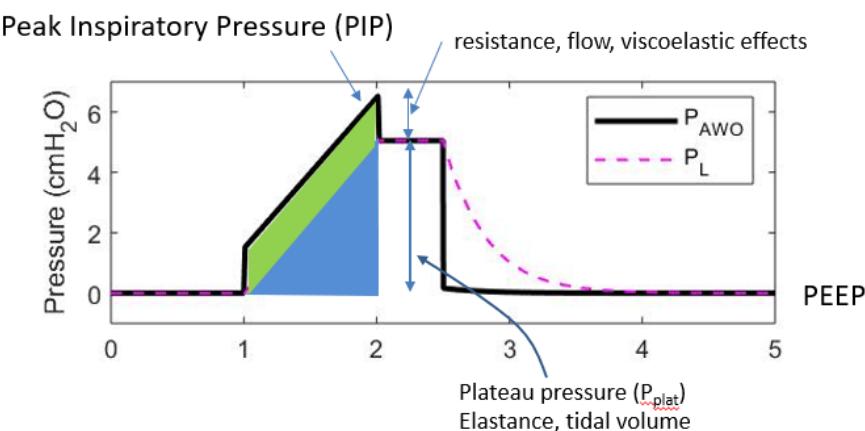


Figure 9-9 Ventilator pressure during a constant flow mechanical ventilation. The green area is the resistive component, the blue triangle is the elastic component of the pressure wave. Several parameters are indicated in the figure.

Note that the expiratory pressure can also be controlled, this is the so-called PEEP pressure which stands for Positive End Expiratory pressure. In this example PEEP is zero. The static lung-chest wall compliance or its inverse the elastance can be estimated from:

	$\text{Elastance} = \frac{1}{\text{Compliance}} = \frac{P_{plat} - P_{PEEP}}{\text{Tidal Volume}}$	9.2-2
--	--	-------

The airway resistance is estimated from the pressure drop at the end of inspiration and begin of the hold period and flow rate Q :

	$\text{Resistance} = \frac{PIP - P_{plat}}{Q}$	9.2-3
--	--	-------

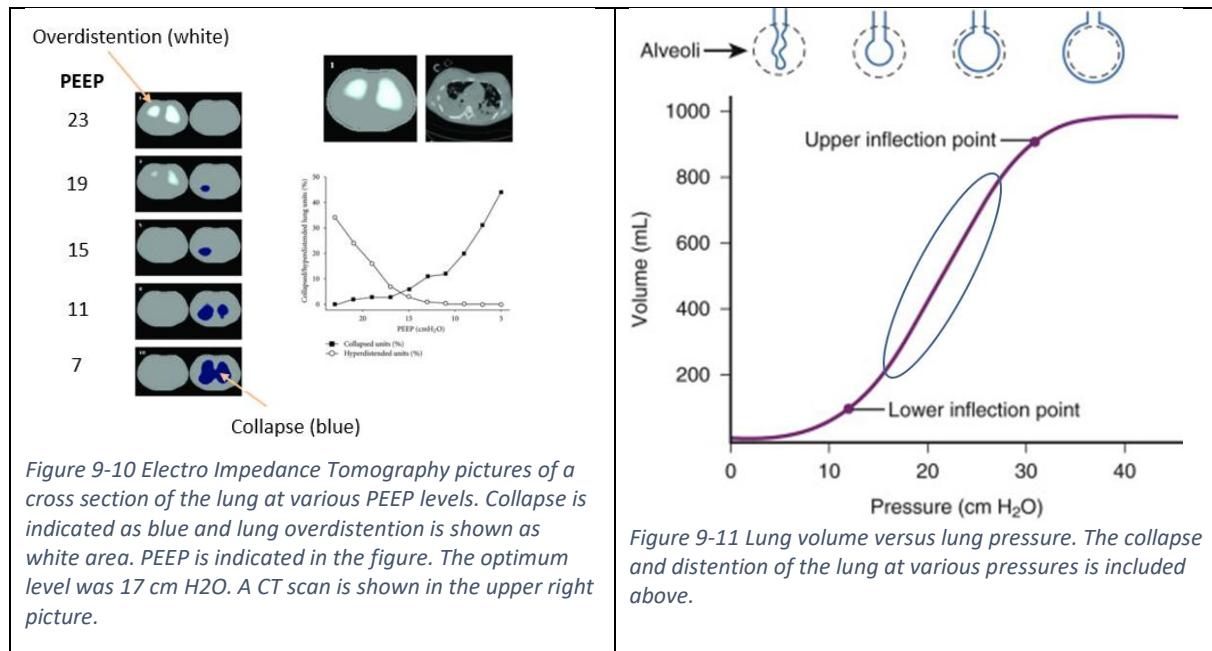
These parameters are of importance for the clinician to optimize the ventilation parameters and to give insight into the status and changes of the patient pathophysiology during treatment. A positive PEEP pressure is often applied to the expiratory system of the ventilator. This is discussed in the next section.

9.2.3 Positive End Expiratory pressure (PEEP)

Patients with respiratory failure have a strongly reduced aerated lung volume. Inflammation triggers a collapse of (a large) part of the lung, lung fluids fill part of the lung and lung ventilation in these areas is not possible. An increase of the aerated volume is needed to improve oxygenation of blood and removal of expired CO₂. For this patient group it is important to apply a positive-end-expiratory-pressure (PEEP) to recruit aerated lung volume. In short PEEP is applied to ventilated patients to reduce alveolar collapse at the end of expiration and improve the ventilation-perfusion ratio. This improves oxygenation of blood. Another benefit is that the airway resistance is reduced.

Furthermore lung injury occurs after repeated collapse and opening of alveoli, this is called atelectasis trauma. The use of PEEP can reduce repeated closing and opening of alveoli. PEEP also reduces the work of breathing when there is patient effort, this is especially important for patients with stiff lungs during spontaneous breathing periods. PEEP reduces energy consumption for spontaneous breathing and reduces the production of large quantities of CO₂. Application of PEEP is important for ARDS, COPD and many other patient groups. PEEP is set at the start of mechanical ventilation, the value depends on the severity of the lung disease. The effect of PEEP on lung volume, atelectasis and lung over-distension is shown in Figure 9-10 for a patient with ARDS. For this patient at low PEEP levels there is a considerable area of the lung that is collapsed (blue area). At high PEEP, collapse disappears but the highest pressures lung overdistention is observed which is harmful (white area). The optimum PEEP level is between the region where lung distention and collapse is observed, in this case 17 cmH₂O.

The influence of lung pressure on the lung volume of a patient with respiratory failure is shown in Figure 9-11. At low pressure, lung collapse is prominent and the V-P relation is convex. For pressures above the lower inflection point the lung volume increases rapidly and aerated lung volume increases rapidly, lung compliance is increased. At higher lung pressure the lung becomes overdistended and very stiff, increases in pressure do not add much extra volume, the lung compliance is reduced, the pressure drop over the lung tissue increases and lung damage and trauma is likely. It is recommended to apply PEEP at a level above the lower inflection point and limit maximum airway pressure below the upper inflection point. The application of PEEP also has drawbacks, it increases the maximum lung pressure but also has an impact on cardiac output. In the next section the impact of positive pressure ventilation on lung pathophysiology and lung injury is discussed.



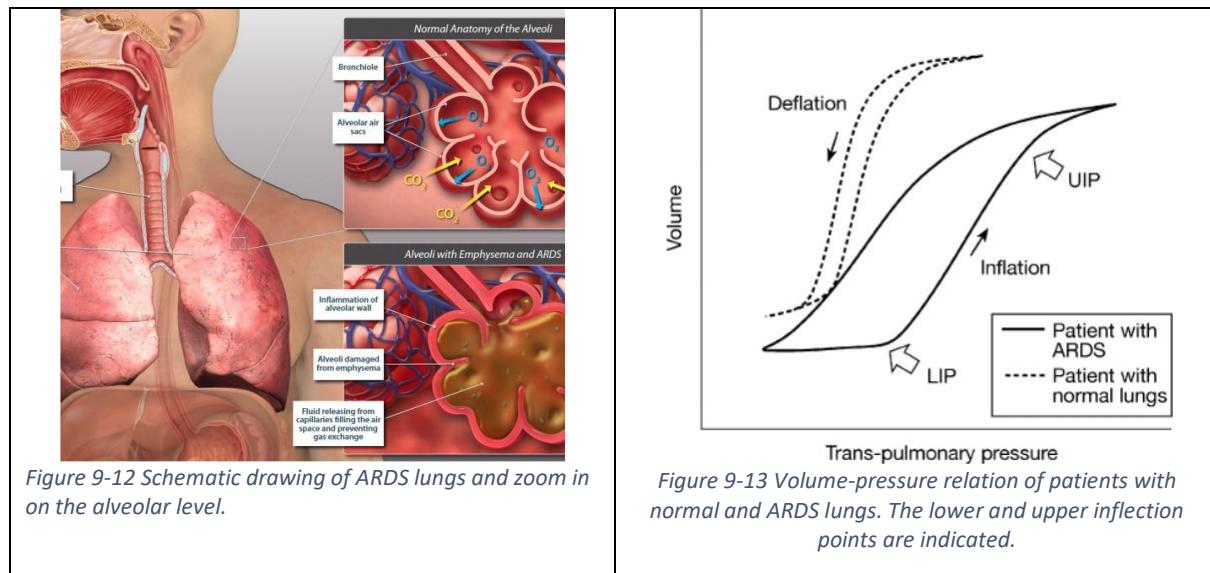
9.2.4 Ventilator Induced Lung Injury (VILI)

The application of relatively large positive pressures on the lung is potentially hazardous. For the healthy lung of for instance a scuba diver, transmural pressure, i.e. the transpulmonary pressure, should not exceed 100 cmH₂O, lethal lung injury is possible at larger pressures. Note that the relevant pressure for lung injury is the difference in pressure over the alveolar wall (i.e. the transpulmonary pressure ($P_L - P_{IP}$)). For the diseased and inflamed lungs of patients with respiratory failure, the maximum pressure that should be applied during PPV is much smaller. However relatively large pressures are needed to increase the aerated lung volume and open up the airways. These high pressures may exceed the threshold for lung damage. Hence there is a complex trade-off between sufficient lung ventilation and ventilator induced lung injury (VILI). Furthermore PPV has a large impact on hemodynamics, high lung pressures reduce cardiac output, this is the topic of the next section. A summary of VILI and impact of PPV on hemodynamics is given below. The focus is on patients with ARDS and COVID like respiratory failure.

ARDS (Acute Respiratory Distress Syndrome) is a rapidly progressive disease with a high mortality rate. Inflammation of the lung causes fluid buildup in the lung and these fluids fill the tiny alveoli in the lung (see Figure 9-12). This lung fluids prevent ventilation and gas exchange with blood of large parts of the lung and causes hypoxemia and hypercarbia of the tissues and organs in the body. This is a life threatening condition and mechanical ventilation is a lifesaving therapy. COVID-19 causes similar lung inflammation and mechanical ventilation is for a large part similar to that of ARDS.

The volume-transmural pressure relations of a normal and an ARDS lung are shown in Figure 9-13. The normal lung shows a small hysteresis in a P-V plot, a rapid increase in lung volume, its lung compliance is large and lower pressures can be used during PPV for a desired tidal volume. The ARDS lung is very different. During inflation the lung volume does not increase initially, followed by a steep increase. At larger pressures the lung is overdistended. The deflation trace is different, there is a very large hysteresis. This anomalous behavior is caused by opening of collapsed parts of the lung. This requires a certain threshold pressure. This increase of aerated lung volume with increasing lung pressure is called recruiting of lung volume. The difference in the deflation trace is related to the

opening of a part of the lung, air can flow out relatively easily. This illustrates the importance of PEEP, this pressure should be larger than that of the lower inflection point.



CT scans of the lung are shown in Figure 9-14. There are relatively dark areas, these are aerated parts of the lung. They are indicated with the green lines. The large “white” part of lung is described as white diffuse ground glass, opacification. This part of the lung consists of the collapsed and fluid filled alveoli and airways. CT imaging of ARDS patients has shown that the aerated part of the lung has similar mechanical properties as that of the healthy lung, i.e. has similar compliance. Air flow into the lung is mainly towards this small volume. Typical volumes of the aerated parts are in the order of 500 ml, i.e. the size of a pediatric lung. The remaining part of the lung is stiff and there is very small lung ventilation in this part. This model of the ARDS lung is the baby-lung model [3]. Note that the “baby lung” is not an anatomical structure, the fluid can accumulate in other parts of the lung when the posture is changed. The baby lung volume is largest when the patient is in the prone position. A second advantage of this position is that there are more blood vessels in the aerated region, the ventilation-perfusion ratio is improved. The function of PPV is to increase the baby lung volume, ventilate this volume and buy time for further treatment and recovery of the patient. Finally the fraction of oxygen (FiO_2) in the inspired air is increased to improve oxygenation of blood. Values as high as 80% to 100% are used in the case of severe ARDS.

Overventilation of the small “baby lung” can cause severe lung injury in this part of the lung. Increases in mortality due to ventilation with too large tidal volume have been reported in several large multi-center clinical studies. Both high pressure and high tidal volume correlate with increased lung injury. Gattitoni and Presento [3] explained this type of lung injury using a simple stress-strain relation of the baby-lung. This relation is shown in the figure below.

$$\text{stress } \sigma = \text{lung pressure} - \text{pleural pressure} = E \cdot \frac{\text{tidal volume}}{\text{FRC}} = \text{strain} \quad 9.2-4$$

E is the Young modulus of the lung tissue, it is proportional to the lung elastance, the inverse of the lung compliance. FRC is the free residual capacity of the baby lung, note that this FRC is not known, it is typically a few hundred milliliter. The maximum lung pressure can be measured, it is the plateau pressure (no airflow condition). The pleural pressure is unknown, it can be estimated when an esophageal catheter is used.

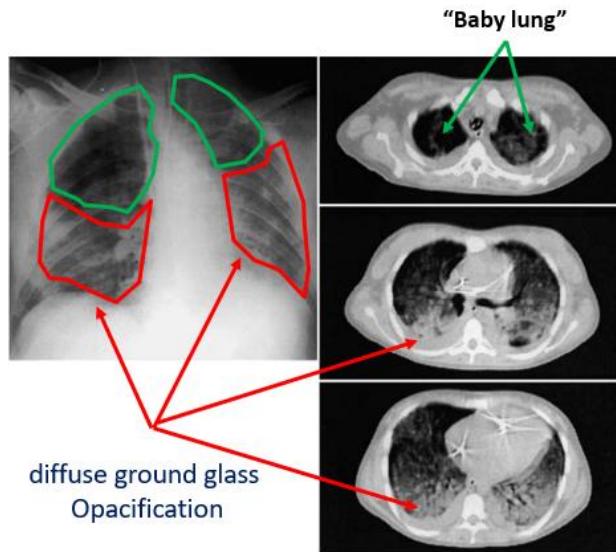


Figure 9-14 Chest X-Ray image and CT scans of ARDS lungs. The dark areas are aerated parts of the lung. The white diffuse ground glass area of the lung are collapsed fluid filled parts of the lung. (From Gattitoni, Presento , Applied Physiology in Intensive Care Medicine, 2009)

This catheter is not often used as it poses extra work to the clinician and increases hazard for the patient. The tidal volume is known, it is measured by integrating the flow during the expiration phase. The driving pressure is the difference between lung and pleural pressure. The strain is the relative increase in aerated lung volume. For a normal lung the FRC is roughly 2 liter, for a tidal volume of 0.5 liter the increase in stress and strain 25% (i.e. $TV/FRC = 0.5/2$). For an ARDS lung the FRC is the baby lung volume and for example it is 500 ml. The relative increase in stress and strain is equal to 100%. This is a large increase and it can damage the fragile lung tissue, this leads to a further reduction in FRC and causes a positive feedback loop towards more lung injury. This type of lung injury is related to increased strain of the tissue and is called barotrauma and volutrauma. The equation shows that in the baby-lung model they are related. Both too high pressure and too large tidal volume can cause VILI. Note again that it is the driving pressure (i.e. pressure drop over the alveolar wall) that is important and not the absolute lung pressure. In practice the pleural pressure is unknown and the airway pressure is used as a surrogate for driving pressure. Large multicenter studies have compared low- and large tidal volume ventilation and authors concluded that small tidal volume was beneficial and correlated with reduced mortality. In the clinic the tidal volume is normalized on the predicted body weight (PBW) with is the length of a patient minus 100 cm. The recommended tidal volume is 6 ml/kg PBW. Often default settings of tidal volume are in the range of 8 to 12 ml/kg PBW. Based on these studies it is recommended to use the following lung protective ventilation settings (data used from the Erasmus hospital in Rotterdam (see table below)).

	Mild ARDS	Moderate ARDS	Severe ARDS
Tidal volume	6 ml/kg	6 ml/kg	4-6 ml/kg
Driving pressure	< 25 cmH ₂ O	< 25 cmH ₂ O	< 25 cmH ₂ O
PEEP	5-10 cm H ₂ O	10-15 cm H ₂ O	20 -25 cm H ₂ O
FiO ₂ (fraction O ₂)	0.20 – 0.40	0.40 – 0.60	0.60 – 0.80
SpO ₂	90 % \geq SpO ₂ \geq 95 %	90 % \geq SpO ₂ \geq 95 %	88 % \geq SpO ₂ \geq 93 %
pH	\geq 7.1	\geq 7.1	\geq 7.1

When there is no esophageal catheter sensing the driving pressure is not known, in this case it is recommended [ref: ARDS net] to limit pressure a tidal volumes according to the following table.

Tidal volume: 6 ml/kg PBW, adjust to plateau pressure	
Plateau pressure: < 30 cmH ₂ O	
Respiration rate: 6 – 35 breaths/min	
I:E ratio: 1:1 to 1:3	
Arterial oxygen pressure: 55 – 80 mmHg	
SpO ₂ : 88% - 95%	
PEEP: 5 – 24 mmHg,	

The above guidelines are important for reduction of VILI but they are not personalized for the specific patient yet and may not be optimal for oxygenation of blood. Furthermore this optimum is highly patient and time dependent. In the “baby lung model” a very important parameter is the FRC of this small “normal lung”. When this FRC is relatively large a larger tidal volume than 6ml/kg is better and should be used. Hence a small tidal volume should be the starting point and ventilation parameters should be further optimized using the observed monitoring data and blood gas data as a guide line. To reduce VILI a small tidal volume is important, therefore control of the tidal volume is very important.

For the acute lung injury and ARDS, COVID patients long duration (days up to many weeks) and relatively high average lung pressures (PEEP up to 20 cmH₂O) are used. The high lung pressure has a severe impact on hemodynamics in the thorax and can result in a large lowering of cardiac output. This is summarized briefly in the following section.

9.2.5 Pro-Con Analysis of PC and VC controlled Modes of Ventilation

The two most used mandatory ventilation techniques were discussed in the previous sections. In short in pressure controlled ventilation the flow rate is the dependent parameter and the tidal volume depends on inspiratory pressure, inspiration time, parasitic tube impedances, and patient physiologic parameters like lung-chest wall compliance and airway resistance. The tidal volume is not controlled and this can result in too low tidal volumes when maximum pressure and inspiratory time is set too low. For the flow-volume control technique the airway and lung pressure is the dependent parameter. The tidal volume is controlled. For a given inspiratory time and tidal volume the airway pressure can be too high and lung injury may be the result. These are in short the main drawback described in many articles and books. Note that in modern computer controlled ventilation machines there are many options and modes that resolve the issues described above. For instance there are so-called volume targeted pressure controlled modes.

There have been many studies of the pros and cons of these PVV technologies, they were summarized in a review article [1]. In short the following effects were reported.

PRO - PCV

- Increased mean airway pressure can improve oxygenation
- Increased duration of alveolar recruitment
- Pressure limited ventilation protects against barotrauma

CON - PCV

- Tidal volume is dependent on lung-compliance and it may vary substantially over the course of mechanical ventilation
- Uncontrolled volume may result in volutrauma
- High initial airway flow may cause problems

PRO - FVCV

- Guaranteed tidal volumes (over time)
- The minute volume remains stable with changing pulmonary characteristics.
- Lower initial flow rate avoids a high resistance-related early pressure peak
- Protects against volutrauma

CON - FVCV

- The mean airway pressure is lower with volume control ventilation
- Recruitment may be poorer in lung units with poor compliance.
- The maximum airway pressure may exceed critical limits for ventilator induced lung injury

In the past the flow-volume control method was used most frequently (~60% of the cases). Recently the pressure controlled system is used more frequently, mainly because of claims regarding reduced lung injury. The most important is of course the clinical outcome expressed in mortality rate. In [1] the conclusions are: *"The two modes have different working principles but clinical available data do not suggest any difference in the outcomes. We included all identified trials, enhancing generalizability, and attempted to include only sufficient quality physiologic studies. However, included trials were small and varied considerably in quality. These data should help to open the choice of ventilation of patients with acute respiratory failure."* Furthermore as mentioned before there many other modes of ventilation that combine the advantages of pressure and flow controlled ventilation. In short more work is needed.

9.3 Ventilator Systems

Ventilator gases have to fulfill many requirements with respect to purity. Sometimes expensive anesthesia gases are used. These gasses have to be removed from the patient area as they can be toxic for the clinicians. In short the gases are expensive and pose risk to the users, Two systems are used, one in which the expensive gases are used one time, another system reuses part of the expiratory gases.

9.3.1 Open Ventilator System

A diagram of an open system is shown in Figure 9-15. The fresh gas mixture composed in the inlet system flows into the ventilator where pressure levels and flow rates can be set. The gas is pressurized via a pneumatic system and the gas mixture flows via the inspiratory valve to the breathing system towards an y-piece and the endotracheal tube into the trachea of the patient. The inspiratory and expiratory valves guarantee unidirectional flow and reduced dead space. The bellow in the inlet system is a passive element and has several functions. The downward movement of the bellow shows that gas flows into the patient and the upward movement shows that fresh gas flows into the bellows during the expiration phase. The top of the bellows corresponds with a volume and the anesthetist has an independent and reliable indication of the tidal volume. The bellow is also a buffer volume of inspired gas. A relief valve is placed in case of malfunction and when rapid action is needed. In modern systems the bellows is absent and replaced by servo controlled pneumatic elements and modern sensing technology, The sensing system is very reliable and waves and pressures are shown on the monitor display (a typical modern system is described later).

The inspiratory and expiratory valves can either be of the passive type or electronically controlled. When the inspiratory valve is opened the fresh gas mixture flows into the lung. Either a constant

pressure or constant gas flow mode with selected parameters can be set by the user. During the expiration cycle the inspiratory valve is closed, the expiratory valve is opened and expiratory gas flows towards a scavenging system.

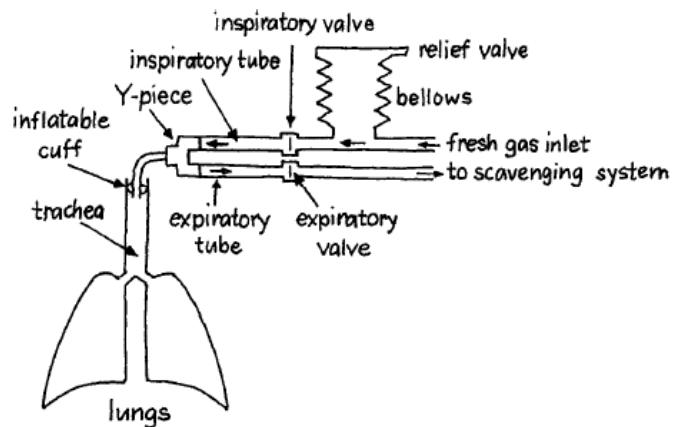


Figure 9-15 Diagram of an open system, the expiratory gas is not reused.

The expiration phase is passive, air will flow from the lung via the expiratory tube to the scavenger as the gas mixture may contain toxic gases and bacteria and virus particles. These contaminations are not allowed in the room ambient. The open ventilator is relatively simple but expensive anesthesia gases and oxygen are disposed in the scavenger system. This leads to high costs, this can be resolved by reuse of these gases in a closed system. This is described in the next section.

9.3.2 Circle System

In a closed system or circle system the expiratory gas is cleaned, humidified and reused. The consumed oxygen and anesthesia gases are added to the processed expiratory gas mixture and this gas mixture is used during the inspiratory phase of a following ventilation. A schematic drawing of a circle system is shown in Figure 9-16. The consumption of fresh gas is reduced with a factor of three. The gas mixture is compressed at the outlet and flows via bacterial filters and a CO₂ absorber back to the inspiratory branch of the ventilator. The valves can be passive unidirectional elements. Note that oxygen and anesthesia gases need to be replenished. This requires a more complex system with additional measurement and control systems. Most modern ventilators are of the circle type.

Note that such systems are of particular interest for the operating theater where the reuse of the expensive anesthesia gases and high purity nitrogen and oxygen gases leads to significant cost savings.

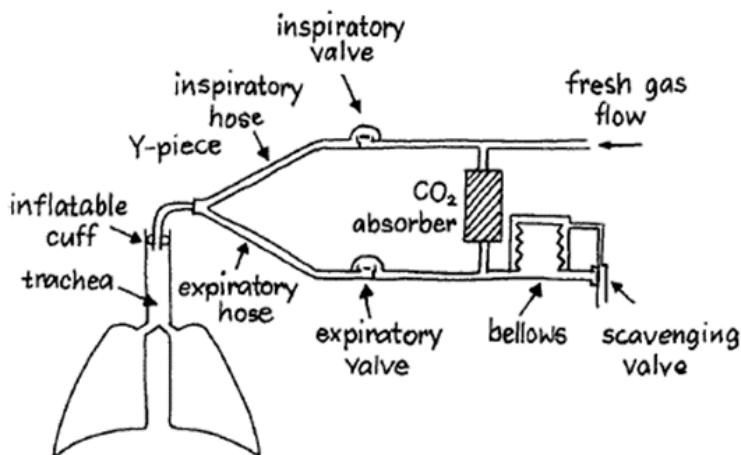


Figure 9-16 Circle ventilation system where the expired gas mixture is partly reused.

In the following section the gas inlet system is discussed

9.3.3 Gas Inlet System

The gas mixture that is inspired by the patient must be of a known and controlled composition. The oxygen partial pressure and anesthesia gases partial pressures can be set by the anesthetist. The oxygen fraction must be measured and some gases are toxic in high concentrations. The concentration must be measured and controlled.

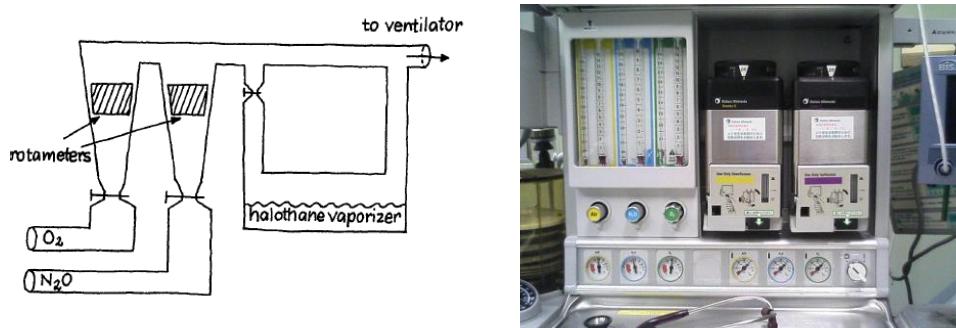


Figure 9-17 Diagram of an inlet system of an anesthesia ventilator. The gas flow to the patient is a mixture of fresh air, pure oxygen and anesthesia gases (N₂O, halothanes).

A diagram of an inlet system of an anesthesia machine is shown in Figure 9-17. It consists of a number of regulators and rotameter flow meters that together determine flow and composition of the gas that flows towards the patient. Part of this flow passes a halothane vaporizer which contains a fluid. The gas above the fluid is added to the gas mixture and is used in the induction and maintenance of anesthesia. Most used halothanes are isoflurane and sevoflurane (simple organic molecules with several Fluor atoms). Typically one percent of the gas mixture that flows to the patient consists of a halothane gas. These anesthesia gases can have a profound interaction with the cardio-vascular system, cause vasodilation and reduce systemic resistance, too high concentrations can lead to a fatal shock condition. Hence concentrations must be controlled and guaranteed. Concentrations of anesthesia gases are monitored and controlled using modern sensing technologies. The gas that flows to the patient is humidified and excess concentrations of CO₂ are removed.

9.3.4 Breathing Set

The gas mixture flows to and from the patient via flexible tubes, the inspiratory and expiratory tubes are separated via a Y-piece and valves (see Figure 9-18). The unidirectional valves are placed in the ventilator. Note that the tubing adds extra dead space to the physiologic dead space and as mentioned before the parasitic resistances cannot be neglected. This extra volume after the one way valves is made as small as possible.

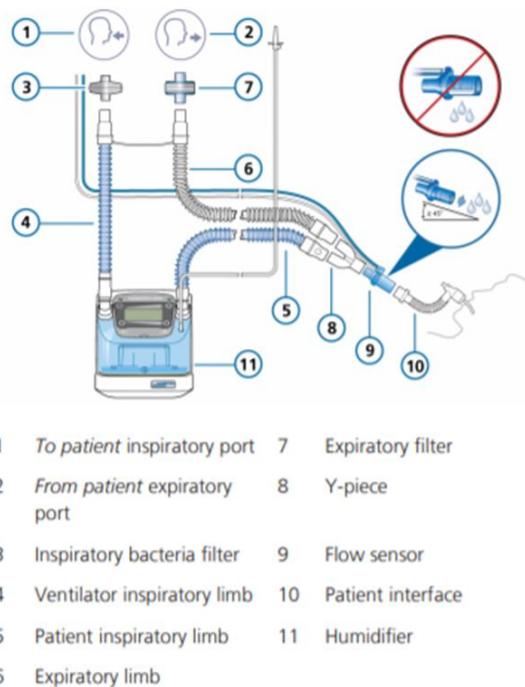


Figure 9-18 Breathing circuit for inspiratory and expiratory gas flow.

Near the ventilator ports ((1),(2)) and before the endotracheal tube (8) filters are placed to remove bacteria and viruses from the airflow to and from the ventilator. This protects the patient and reduces the risk of contamination of the ventilator machine. Note that ventilator acquired pneumonia is an important source of hospital infections. The best location for a (volumetric) capnography sensor and flow and pressure sensors is directly before the connection to the patient (location 9). The Philips Capnostat combines these functions in a single compact unit. A humidifier (11) is added to warm and moisten the inspired air.

The measured flow resistance of the inspiratory, expiratory and endo-tracheal tubes is shown in Figure 9-19 and Figure 9-20. The flow resistance is proportional to the flow rate due to turbulent air flow. During mechanical ventilation the maximum flow rate is in the range of 0.5 l/s to 1 l/s. Hence the sum of the maximum tube flow resistances are in the range of 1.5 to 10 cmH₂O.s/l. The airway resistance of a patient with normal lungs and airways is in the range between 1 and 3 cmH₂O.s/l. Note that the parasitic resistances of the breathing circuit and endotracheal tube are much larger than that of the patient airways for a patient with normal lung function. This has a large impact on the work of breathing when the patient breaths through the tubes and when the ventilator does not support the patient breath. Furthermore these parasitic resistances have a large influence on the measured pressure and flow waveforms near the Y-piece. The clinician uses the proximal pressure waveform to monitor PPV. Finally the tubes also have compliance, for a 180 cm length tube the compliance is in the range between 0.15 ml/cmH₂O and 0.3 ml/cmH₂O. For a driving pressure of 20 cmH₂O this corresponds to a volume change of 30 ml to 60 ml, this volume change has to be subtracted from the measured tidal volume (~500 ml) obtained by integration of the airway flow.

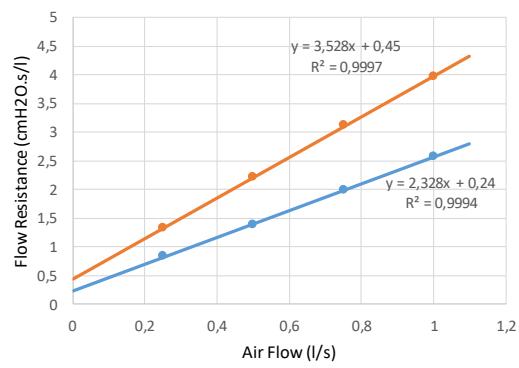


Figure 9-19 Flow resistance versus flow rate of coaxial inspiratory and expiratory tubes (length 180 cm). The resistance increases with flow rate due to turbulence and follows Rohrer's equation $R = K_0 + K_1 \cdot Q$.

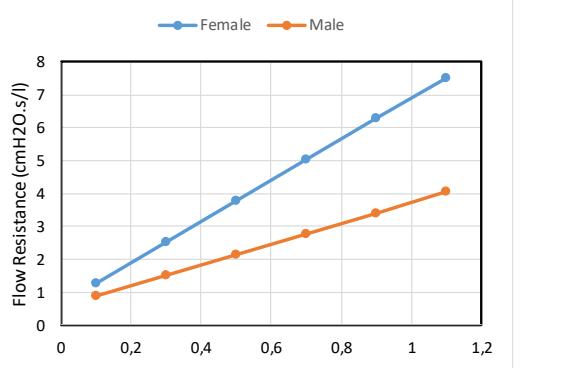


Figure 9-20 Flow resistance of endotracheal tubes with internal diameter of 7 mm (female) and 8 mm (male) versus flow rate. (ref [2]). Resistance corrected for sensing tube.

9.4 Modern PPV Ventilator

There are several architectures and methods used in modern ventilators. A Hamilton G5 system is described in this section. The approach is widely in use. Modern high-end ventilation machines are complex computer controlled pneumatic devices with many sensing and control functions. The essential physiologic monitoring functions are integrated in the ventilation machine. These machines become more intelligent and support the clinician to optimize the ventilation process, minimize barotrauma and assist in the weaning process. An example of a state of the art ventilator is the Hamilton G5 (Figure 9-21)Figure 9-21.

The device has integrated monitoring of the airway pressure, gas flow, volume, pulse oximetry and volumetric capnography. The volumetric capnography method includes the P_{CO_2} waveform, ETCO₂ and additional parameters from the CO₂-volume waveform and the minute volume of CO₂ expired gas volume (VCO₂) and is useful for optimizing ventilation parameters. VCO₂ is an important indicator of the quality of ventilation and metabolic activity and can be used to estimate dead space volumes. A touch screen with the user interface is used for display of data, display of device and physiologic information and control of the device. The main components and functions are described next.

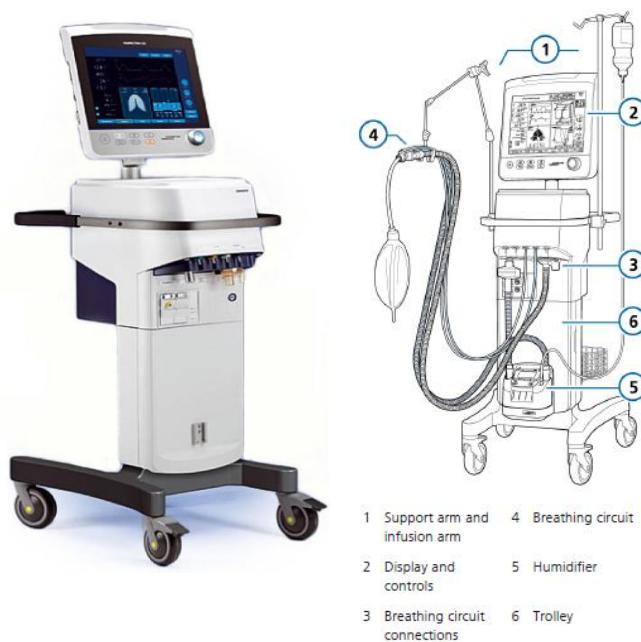


Figure 9-21 Hamilton G5 ventilator and main components.

9.4.1 Principal Gas flow in the G5 Ventilator

A diagram with the principal gas flow is shown in *Figure 9-22*. Compressed air and oxygen from external sources are connected to the device. The inlet pressures are in the range of 4 to 5 bar. Gas flow is regulated at three locations. The first location is point C, the mixer block and controls the pressure in point D, a large tank. Air and oxygen flow via two switches to the mixer block (C) where the desired gas composition (i.e. the oxygen fraction FiO_2) and reduced pressure is obtained. The mixed gas flows to a tank (D) with an overpressure between 200 mbar and 340 mbar. This tank functions as a buffer volume of gas of a precise controlled composition and the higher tank pressure is used to drive gas flow in the remainder of the system. The second location is the inspiratory valve. Gas flows from the tank to the inspiratory valve (E) where flow and inspiratory pressure are measured and regulated and gas flows into the inspiratory tube of the breathing circuit. The third control point is the expiratory valve (F). Expired gas flows to the expiratory valve where the expiratory pressure is controlled (i.e. PEEP).

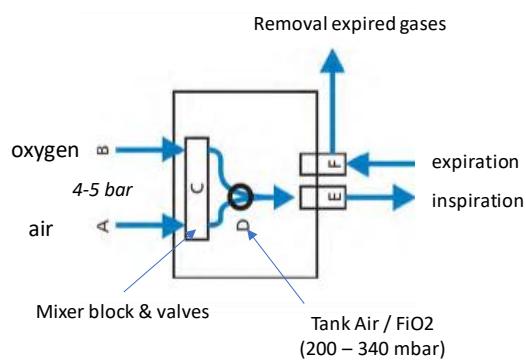
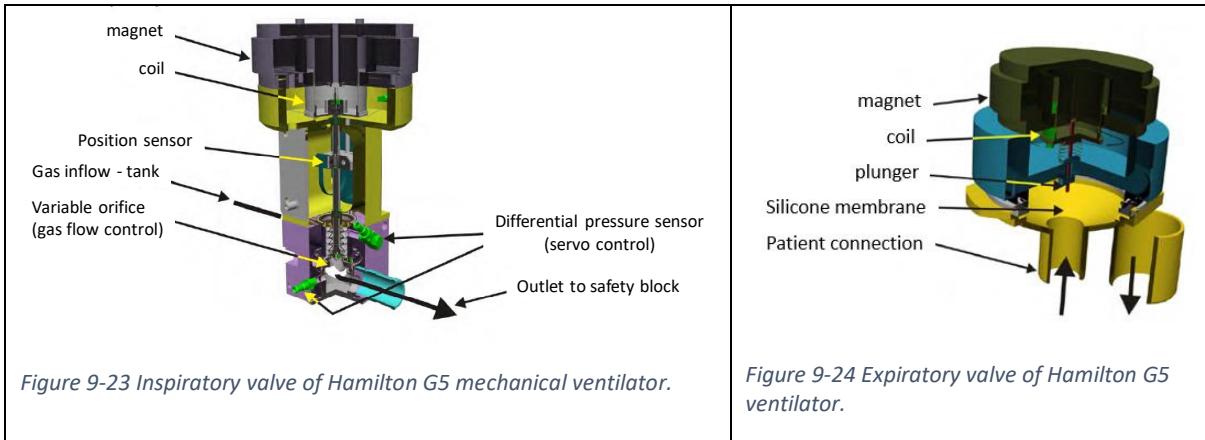


Figure 9-22 Principle gas flow in the Hamilton G5 patient ventilator.

The inspiratory (E) and expiratory (F) valves are shown in *Figure 9-23* and *Figure 9-24*.



The inlet port of the inspiratory valve is connected to the tank, gas flows through a variable orifice, its size or flow resistance can be controlled by an electro-magnetic actuator. The flow is calculated from the signal of a differential pressure sensor and the known position of the plunger. The flow or pressure are servo controlled towards the set flow rate (VC mode) or set pressure (PC mode). A maximum flow rate can be set during the PC mode. In a volume targeted mode the tidal volume can be limited to a range set by the user. The lower chamber pressure is the outlet pressure. The outlet port is connected to a safety block where a mechanical valve limits over pressure to 110 mbar. Furthermore in case of failure a valve is opened automatically that allows the patient to breath air from the environment. The expiratory valve is a magnetic actuator that applies a specific force to the silicone membrane that corresponds with the set value of PEEP pressure. For a larger expiratory pressure than the set PEEP value the valve remains open, for a smaller pressure the valve closes. Alternatively the valve can be either open or closed by the actuator. The valve is closed during the inspiration phase. The valve is open in the emergency condition of device failure. An overview of the principle gas-flow of the device is shown in [Figure 9-25](#).

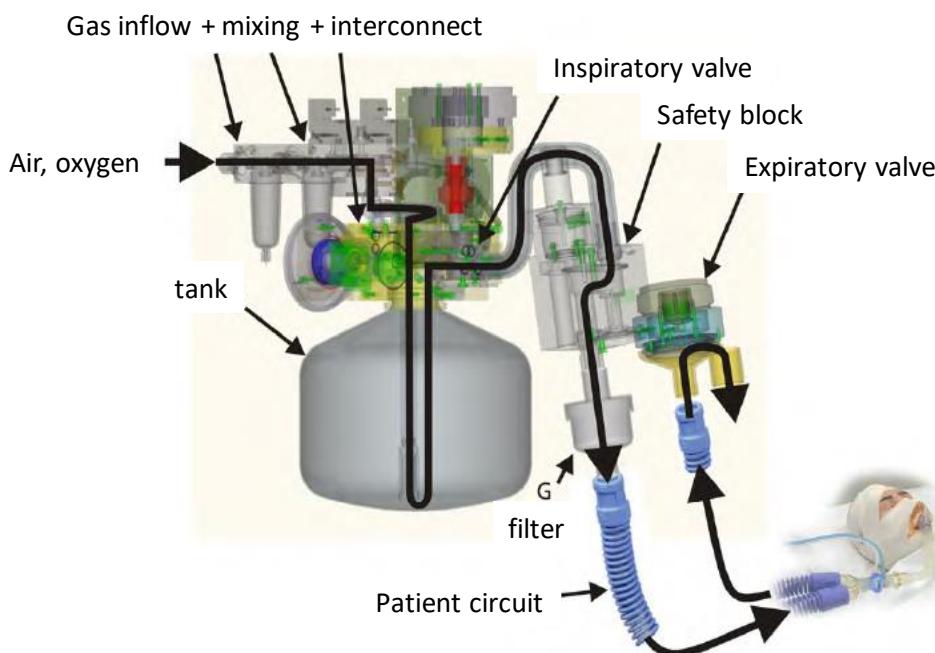


Figure 9-25 Overview of principle gas flow in the Hamilton G5 ventilator.

9.5 Ventilator Modes in a Modern PPV ventilator

The two basic modes of pressure or volume control were discussed in section 9.2. Presently clinicians use many more modes of ventilation. A modern ventilator enables the use of these modes. The options are indicated in the diagram shown in Figure 9-26. The main modes are pressure control (PC) and flow-volume control (VC). The device supports different types of breaths. The first is the continuous mandatory type (CMV) where the ventilator and clinician are in full control of the ventilation rate, inspiratory and expiratory pressures, flow and tidal volume. It is used when the patient is sedated and muscles are relaxed. The ventilator starts and stops the inspiratory phase using the set values for rate, I:E ratio, pressure, flow, tidal volume. When the patient recovers a weak spontaneous effort is present. This must be stimulated to promote recovery and shorten time on the ventilator. In the spontaneous mode the patient has a continuous breathing effort but is too weak to breath without ventilator support (CSV). The ventilator detects the patient effort and supports the breath. The patient determines the start of the respiration and the respiration rate. The ventilator estimates the end of inspiration. Finally there is a mixed mode, the intermittent mode (IMV). There is a fixed mandatory breath rate, between mandatory breaths spontaneous efforts are detected and supported. Finally ventilator breaths can be synchronized (S) with patient effort. The ventilator detects patient effort and supports patient breathing, i.e. patient and ventilator support are synchronous. Note that in the mandatory mode the patient may trigger the breath but the ventilator determines the cycling off time and breathing rate. Hence there are many modes (> 10) and breath types that need to be supported by the mechanical ventilator. For instance PC-CMV: pressure controlled continuous mechanical ventilation, VC-IMV: volume controlled intermittent mechanical ventilation, PC-CSV: pressure controlled continuous spontaneous ventilation, SIMV: synchronized intermittent mechanical ventilation. Finally there are other modes of the shape of waveform (ramp, shape, CPAP, BiPAP, neonates, pediatric) that are supported. Non-invasive ventilation modes are important and must be supported (mask, helmets). The detection of spontaneous patient breaths is described next.

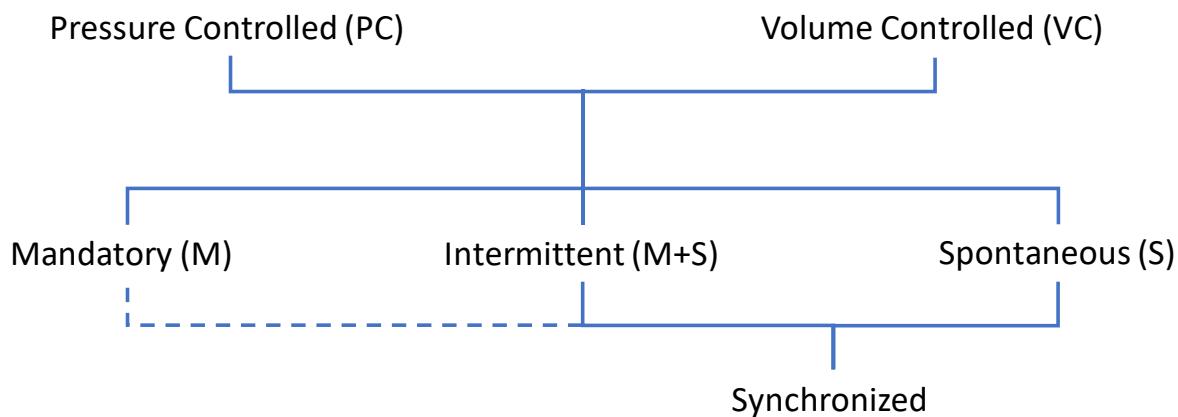


Figure 9-26 Main ventilator mode and supported breath types.

9.5.1 Synchronization of Patient effort and Support Ventilation

Spontaneous breathing is a sign of recovery and this must be promoted. However the patient is still too weak to breath independently and the breathing effort of the patient must be supported. Note that in a spontaneous breathing effort the lung pressure decreases below the set PEEP pressure and (positive) airflow from the breathing set to the lung starts. The airway opening pressure decreases below PEEP, the expiratory valve closes and airflow towards the patient increases and eventually changes sign (see Figure 9-27 and Figure 9-28). This drop in airway pressure and/or increase in

inspiratory flow rate near the patient can be detected by the machine and in a support mode it can trigger opening of the inspiratory valve. It is called pressure triggering when the drop in pressure exceeds a certain user set threshold and flow triggering when the flow exceeds a certain value. Typical pressure threshold values are 1 to 2 cmH₂O below PEEP. When flow is used for triggering it is called flow triggering. Typical flow trigger values are +0.02 to +0.1 l/s. A typical delay from start inspiratory effort to start trigger is 0.05 to 0.15 seconds. The duration of patient effort is around one second.

The patient can trigger the breath but stopping or cycling-off of the inspiration period is not controlled by the patient. The breath should end somewhere between maximum flow peak and the end-time of patient effort. This cannot be inferred from the waveform and therefore different methods are used in ventilators. The method that is used most often is cycling-off at a percentage of the maximum inspiratory flow. This is shown in Figure 9-29. The inspiratory pressure is cycled off at a fixed percentage of the maximum inspiratory airflow. A default value is 25%. A drawback of this method that the 25% value may not correspond to the end of patient effort especially for patients with very high airway resistance or very stiff lungs. Furthermore the tidal volume is not controlled. In practice the cycling off percentage must be optimized for the specific patient.

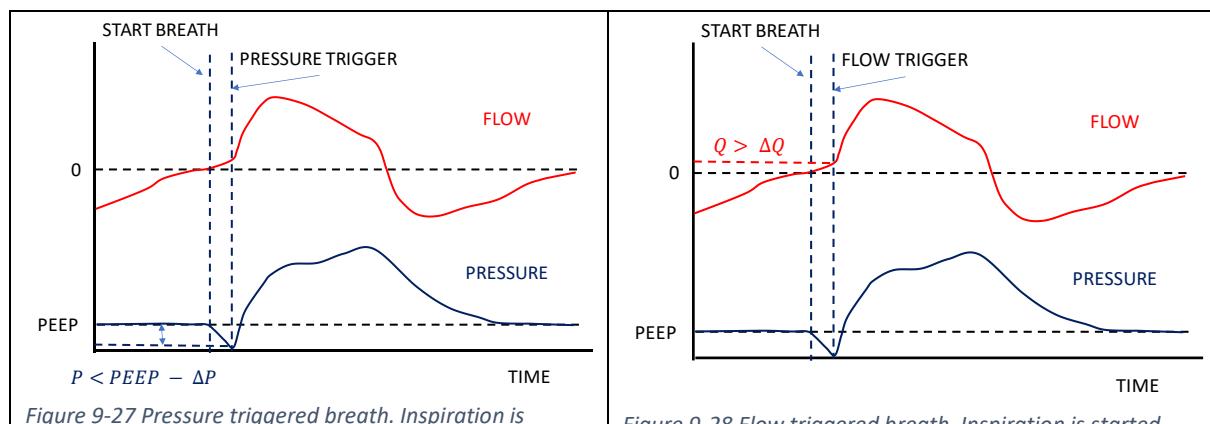


Figure 9-27 Pressure triggered breath. Inspiration is started when the airway pressure drops at a threshold pressure below PEEP.

Figure 9-28 Flow triggered breath. Inspiration is started when the flow in the expiratory phase increases and exceeds a positive flow threshold.

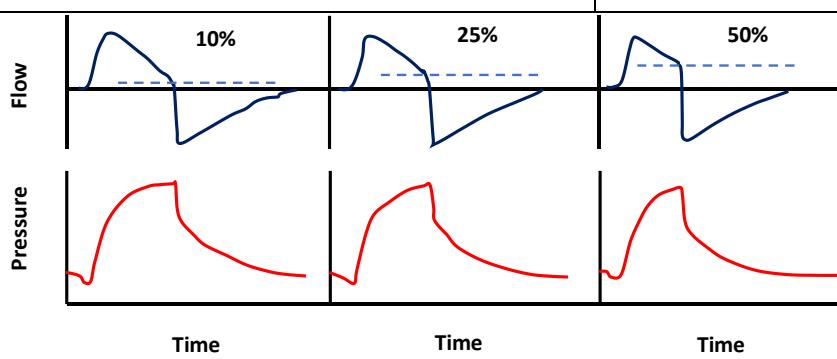


Figure 9-29: Diagram of flow induced cycling-off of a patient pressure triggered breath. The cycling off threshold was 10%, 25% and 50% of the maximum inspiratory flow.

Other methods to start the cycling off are a fixed time or cycle off when a specific tidal volume has been reached.

9.6 Mechanical ventilation process

The different phases during mechanical ventilation in the ICU department are shown in Figure 9-30. When mechanical ventilation is needed the patient is sedated and the endotracheal tube is placed, this is the intubation phase. The indications to start invasive mandatory mechanical ventilation are: poor oxygenation (hypoxemia), hypercapnia (too high CO₂ pressure), dyspnea (shortness of breath), shock state, respiratory failure, too small or a too strong effort (patient causes lung injury!). The patient is sedated, administration of muscle relaxants is started. The mechanical ventilation machine completely controls the breathing process. It can either be pressure or volume controlled and the ventilation parameters tidal volume, ventilation rate, I:E ratio and maximum airway pressure are set by the clinician. This is mandatory mechanical ventilation. It is very difficult to optimize the mandatory ventilation to the patient. The patient demands and patient characteristics are not known, under- and overventilation must be avoided, VILI must be avoided. The impact on hemodynamics must be limited. The patient condition can degrade rapidly and continuous monitoring and optimization of ventilation is necessary. Lung protective ventilation is a must and as mentioned before it is a complex challenge to obtain the optimal patient specific and time dependent ventilator settings. Clinical studies have shown that the mandatory phase must be as short as possible, prolonged mechanical ventilation is harmful for the patient, increases length of stay in the hospital, is correlated with increased mortality and can have a dramatic impact on the quality of life after hospital discharge. Permanent lung damage, damage to the diaphragm and other respiratory muscles and muscle atrophy (shrinking of muscle size) occurs frequently. Muscle injury hinders the patient to start and maintain independent spontaneous breathing after weaning.

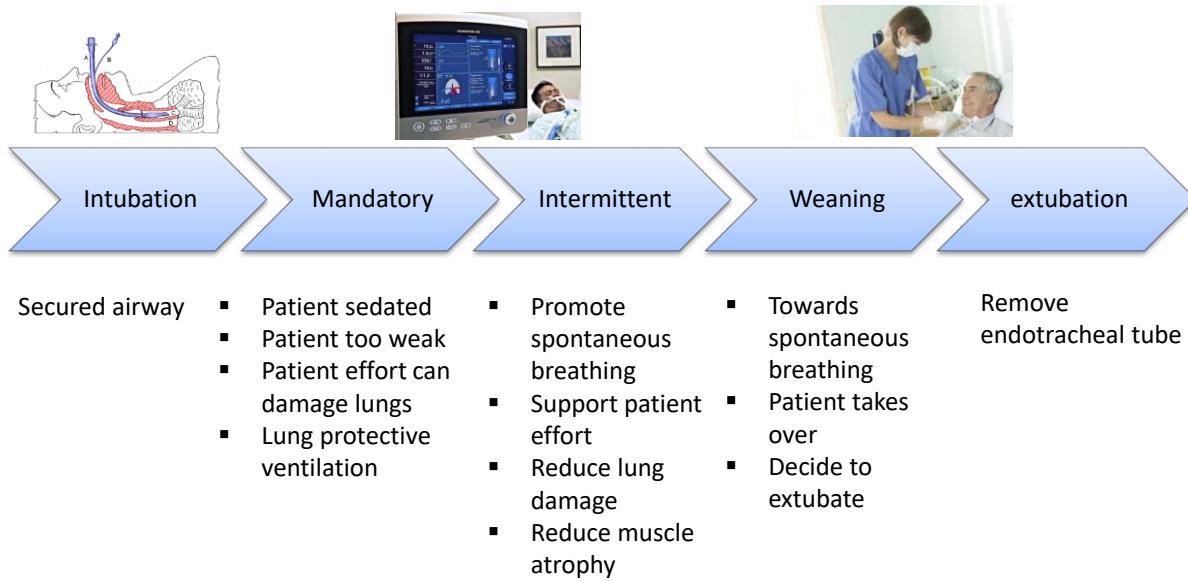


Figure 9-30 Different phases during mechanical ventilation in the ICU environment.

Sedation of the patient is reduced at regular intervals and as soon as the clinicians detect a spontaneous breathing effort of the patient measures are taken to support patient efforts. Initially the patient is not strong enough to breathe independently. Modern ventilators support the spontaneous breaths such that sufficient lung ventilation is achieved and mandatory breaths still occur at fixed times, the desired minute volume is maintained. This is called intermittent ventilation and there are many variants as was described in section 9.5.

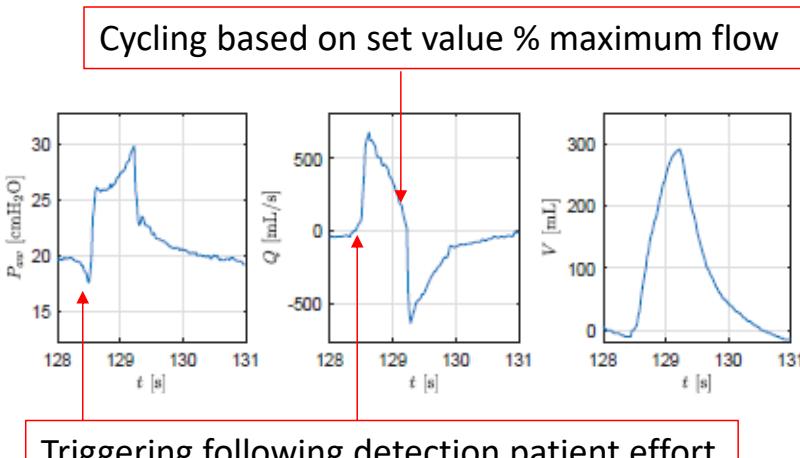


Figure 9-31 Pressure, flow and ventilation volume measured at the airway opening during pressure support ventilation.

Finally the patient effort becomes more regular and stronger and the mandatory breaths are omitted. The spontaneous patient breath is still supported. Again there are many modes of support, a popular method is pressure support ventilation (PSV) of spontaneous breaths. In the PSV mode the patient initiates every breath and the ventilator delivers support with the preset ventilator pressure value at the inspiratory port. With support from the ventilator, the patient regulates their own respiratory rate and tidal volume. A plot of pressure, flow and volume of a normal patient triggered PSV breath (measured at the airway opening) is shown in Figure 9-31. Note that the pressure measured at the airway opening deviates from that measured in the ventilator (i.e. constant inspiratory pressure) because of the large pressure drop over the inspiratory and expiratory tubes. The cycling off time is determined by the time when a set value of the fraction of the maximum expiratory flow is reached (typical value 25%). Unfortunately the triggering and cycling off is not perfect. Studies have shown that desynchronization between patient effort and triggering and cycling off ventilator support is common. A mismatch between the ventilator and patient effort is called an asynchrony. Clinical studies have shown that the number of asynchronies is large (10 percent or more) and asynchronies can increase length of stay in the ICU, may cause lung injury and even increases in mortality have been reported. There are more than 10 types of asynchronies.

To summarize, asynchronies in support and mandatory ventilation are harmful for the patient and the number of asynchronies must be kept low. Automatic identification of asynchronies is needed for clinical studies, training and for smart ventilation machines. In the latter case warnings could be generated and the user could be advised and guided to change the settings of the machine.

9.7 Physiological Closed Loop Ventilation (Smart or Intelligent Ventilation)

Control and optimization of mechanical ventilation of patients with respiratory failure (ARDS, COVID, pneumonia) is a complex process. In the standard care setting the clinician determines the control targets (see Figure 9-32) for the specific patient and changes ventilator settings when needed. The clinician controls the process. Multiple input parameters need to be considered. The physiological measurements include parameters related to oxygenation of blood and carbon dioxide removal (SpO_2 , ETCO_2 , blood gases, lab data) and hemodynamics (blood pressure, cardiac output). Furthermore parameters related to lung mechanics and ventilator settings related to lung injury (tidal volume, driving pressure, PEEP) must be included in the optimization of the therapy. Finally when there is a patient breathing effort this must be taken into account and possibly a support mode of ventilation must be used.

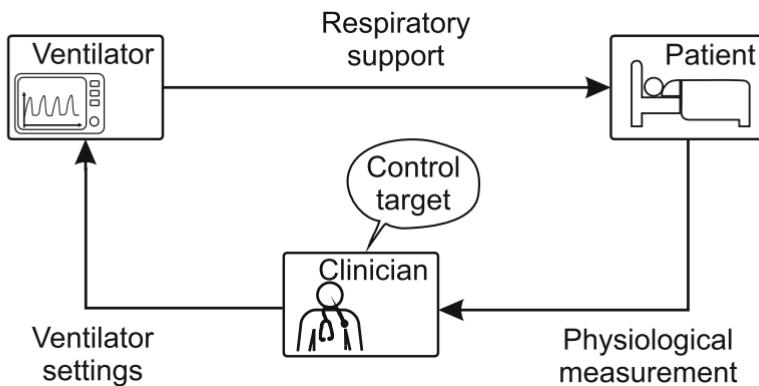


Figure 9-32 Standard clinical care loop, clinician controls the mechanical ventilation (source von Platen et al.).

There are conflicts in ventilator settings when these different type of measurement variables need to be included in the decision process. Patients are often overventilated, i.e. there was too much emphasis on the monitor parameters related to blood oxygenation and CO₂ removal. In most cases the patient disease state varies, becomes more severe, improves in a later phase and regular change of settings are needed. This requires continuous attention of the ICU staff, a high level of expertise is needed to optimize ventilator settings in the ICU. The continuous adjustments are often not possible, there is a shortage of experienced ICU staff. The consequences for the patients can be dire.

There is a need for intelligent ventilation devices that can assist the clinician in real time, improve patient safety, reduce the workload and improve outcome. Due to the progress in electronics and computer technology and increased understanding of physiology and modeling intelligent ventilation smart ventilation techniques are within reach. Such techniques must be safe (reduce patient harm), effective (improve outcomes) and efficient (reduce workload). The leading ventilator manufacturers have proprietary solutions to resolve (part of) the challenges. In this section closed loop ventilation technology is briefly discussed. References to review articles will be given in the appendix for readers that want to have more information on this topic.

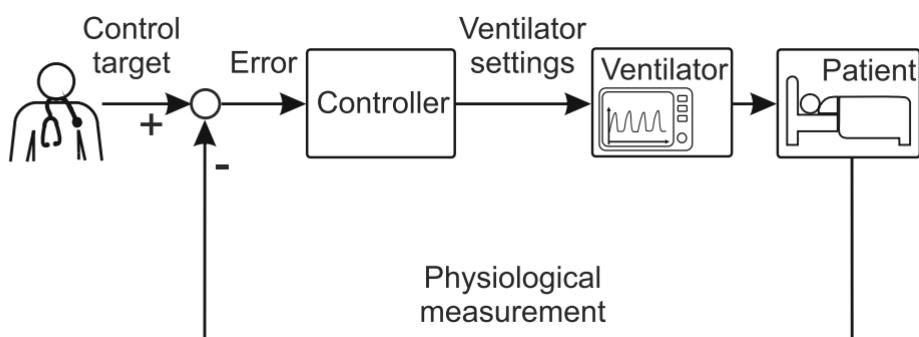


Figure 9-33 Physiological closed-loop mechanical ventilation process. (source von Platen et al.).

A diagram of a closed-loop control process for a mechanical ventilator is shown in Figure 9-33. The clinician determines the personalized control parameters and their range. These values become the set points for the control loop. It is a major challenge to design, develop and validate the control loop. The function is to track the settings and correct for disturbances. The tracking process is the dynamic response of the system until the setpoints are reached. It must be stable. A second demand is robustness for disturbances. Disturbances can be internal (change in disease state, lung mechanics, CO₂ production, spontaneous breathing effort) or external (disconnection, blockage of breathing set tubes, motion artifacts, external therapeutic devices). An example of a system that uses ETCO₂ as the

control variable is shown in Figure 9-34. At time 1 the set point is changed, the minute volume MV is increased to reduce the ETCO₂ level. At time 2 extracorporeal CO₂ removal is stopped, this disturbance is compensated by a second increase in minute volume.

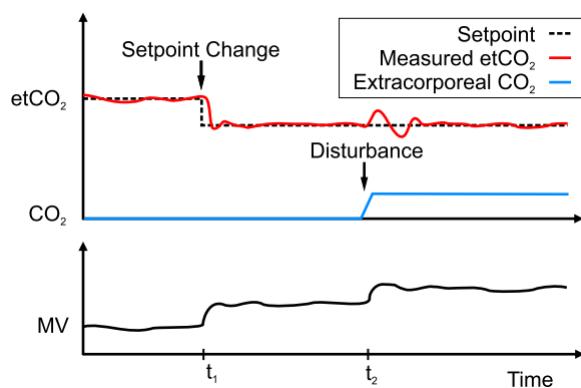


Figure 9-34 Set point tracking and disturbance control in a ETCO₂ closed loop system. At t1 there was a change in setpoint, at t2 there was a disturbance due to an extracorporeal CO₂ removal device. (source von Platen et al.)

In the first systems the control system was targeted on gas exchange, carbon-dioxide concentration was used as the set point variable (ETCO₂, VCO₂). It was necessary to include oxygenation as well and it was added by measuring pulse oximetry (SpO₂). The next step was to include lung mechanics (lung-chest wall compliance, airway resistance), this was important for optimization of ventilator settings (tidal volume, frequency, PEEP) and to reduce ventilator work (i.e. energy deposited in the lung) as well as to improve gas exchange in the lung. A next step was to include lung protective ventilation settings for the tidal volume and driving pressure, this required input of patient parameters like length, weight (PBW) etc. The studies were initially focused on mandatory ventilation, later also support ventilation was studied. When the patient recovers spontaneous breaths are seen but the patient may be too weak to breath independently. In this case the support modes of mechanical ventilation are used. It is even more difficult to optimize ventilator settings in the support modes. The patient effort can be monitored by measuring the decrease in proximal airway pressure at 100ms after start of the breath (P0.1) or by measuring electromyography (EMG) or the neural activity of the respiratory muscles. EMG measures muscle response or electrical activity in response to a nerve's stimulation of the muscle. When the EMG measurement is done on the diaphragm muscle it is called EAdi. A diagram of a modern closed loop controlled system is shown in Figure 9-35.

Targets and ranges for minute volume, maximum inspiratory pressure, PEEP and oxygen (SpO₂) and carbon dioxide (ETCO₂) must be given by the user. Furthermore patient demographic data must be given such that the device can estimate lung mechanics parameters and other parameters that comply with guidelines for reduction of ventilator induced lung injury. Finally when there is a spontaneous breathing effort, it must be detected and the strength of the effort must be estimated by either P0.1 (pressure drop at the airway opening after 100ms of the start of a spontaneous breathing effort) or EAdi (electrical activity of the diaphragm muscle) and the ventilator settings must be adapted to correct for the patient effort to control minute ventilation. There are often different controllers for lung mechanics, oxygenation and carbon dioxide. There are several approaches for closed loop ventilatory control, there are several commercial systems in the market. Two commercially available closed loop systems are discussed in the appendix. These are Adaptive Support Ventilation from Hamilton for use during both mandatory and support ventilation and Neurally Adjusted Ventilatory Assist (NAVA) mainly used during support ventilation.

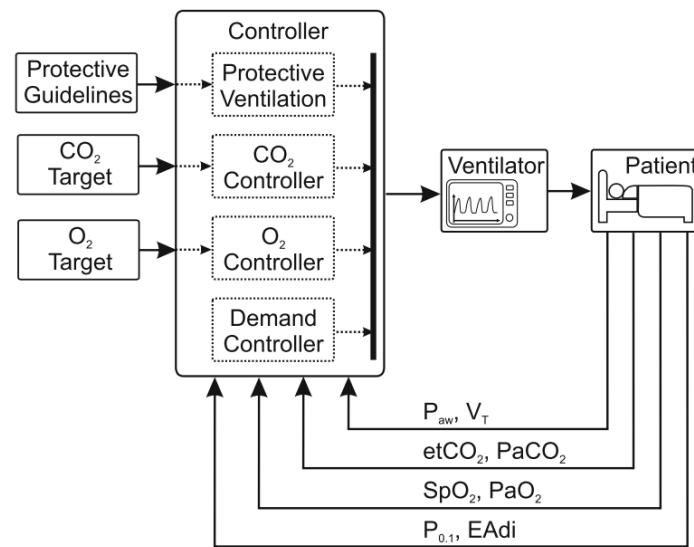


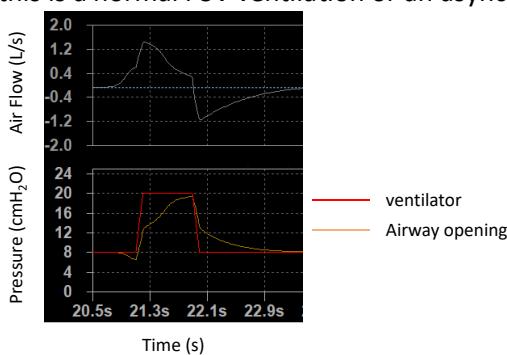
Figure 9-35 Topology of a state of the art fully automatic control-loop. (source von Platen et al.)

9.8 References

1. Nuttapol Rittayamai , MD ; Christina M. Katsios , MD ; François Beloncle , MD ; Jan O. Friedrich , MD , PhD ; Jordi Mancebo , MD ; and Laurent Brochard , MD; CHEST 2015; 148(2): 340 – 355.
2. A. G. FLEVARI, N. MANIATIS, T. E. KREMIOTIS, I. SIEMPOS, A. P. BETROSIAN, C. ROUSSOS, E. DOUZINAS and A. ARMAGANIDIS, Anaesth Intensive Care 2011; 39: 410-417.
3. Luciano Gattinoni and Antonio Pesenti, chapter in M.R. Pinsky et al. (eds.), Applied Physiology in Intensive Care Medicine, DOI: 10.1007/978-3-642-01769-8_54, © Springer-Verlag Berlin Heidelberg 2009.
4. Philip von Platen, Anake Pomprapa, Burkhard Lachmann and Steffen Leonhardt, Critical Care (2020) 24:121.

9.9 Questions Therapeutic Devices

1. Discuss the advantages and disadvantages of ventilating a patient with a ventilator that provides a controlled flow, b) a controlled pressure to the patient's airways.
2. Give a simple model of a closed flow controlled PPV ventilator-patient system that includes the most important elements including parasitic elements.
3. Use this model to sketch the pressure, flow and volume waveforms for a pressure and a flow controlled ventilator. Insert the main parameters.
4. Describe how the open system ventilator functions. When are the valves opened? What is the disadvantage of this system?
5. Describe a method to estimate the airway resistance during PPV. Is this estimate accurate? Explain your reasoning.
6. Explain how the lung compliance can be measured during PPV ventilation.
7. What is the function of a ventilator's inlet combination?
8. Explain why ETCO₂ decreases in an exponential manner with time when the ET tube is accidentally placed in the esophagus.
9. Explain why ETCO₂ obtained from capnography decreases/increases when the minute ventilation increases/decreases?
10. ETCO₂ is assumed to be roughly equal to the CO₂ partial pressure of arterial blood. Is this still the case when the minute ventilation volume is much too large or too small?
11. The clinician aims at an ETCO₂ value of 40 mmHg. Presently the value is 50 mmHg. Explain which settings of the ventilator machine have to be adjusted.
12. What is lung protective ventilation? What settings of the ventilator are needed?
13. What is "assisted ventilation" and how is it realized?
14. What is pressure support ventilation?
15. What are asynchronies during pressure support ventilation?
16. Which monitoring sensor(s) are needed to detect, identify and annotate an asynchrony in a measured time series of PPV ventilation?
17. In the figure below two pressure waveforms (one measured at the ventilator, one at the airway opening) and the flow waveform of a pressure support ventilation (PSV) are shown. Is this a normal PSV ventilation or an asynchrony? Explain your reasoning.



18. Explain the differences between pressure waveform measured at the ventilator and at the airway opening.
19. Explain why PEEP is required when ventilating a patient with ARDS.
20. Describe how the circle system ventilator functions. When the various valves open and close and what determines this?
21. Some commonly occurring problems in artificial ventilation are: disconnection of inspiratory or expiratory hose; leak of inspiratory or expiratory hose; esophageal intubation; bronchial

intubation. Which measurements and/or patient data would be required to discover these problems in an alarm system?

22. What means “closing the loop” for a PPV ventilation technique?

10 In-Hospital Patient Monitoring

In this chapter in-hospital patient monitoring is discussed. In the previous chapters state-of-the-art monitoring devices and techniques were discussed. Historically patient monitoring started in areas where there was the highest risk of mortality. Therefore patient monitoring was first used during surgery and in the intensive care environment. In the ICU there is a high risk of complications and increased probability of mortality. Monitoring in the ICU or during high level surgery is called high-acuity monitoring. This is also called time-critical care. In the first section this classical type of monitoring and use of data in these environments is discussed.

It was perceived that in-hospital mortality would be dominated by these high-acuity departments. However studies from the 1980's up till now have shown that a large fraction of unanticipated deterioration of patients occurs in the hospital wards. The late detection of patient deterioration in the ward leads to a prolonged length of stay in the hospital, unanticipated admission to the ICU and increased mortality. Around 40% of the unanticipated deaths in a hospital occurs in the ward. There is a growing need for systems that tackle this problem in the low-acuity departments in the hospital. So-called rapid response systems have been introduced and there is a lot of evidence that the number of unanticipated deterioration of the health state is reduced. The first requirement in the RRT systems is to generate data that can trigger these systems such that deterioration in the health state can be detected in an early phase. Therefore there is a need to monitor patients in lower acuity environments. However monitoring requirements in the ward differ a lot from those in the ICU and OR. A new generation of monitoring technologies and IT systems is needed. Low-acuity monitoring techniques are discussed in the second section.

Monitoring devices and architectures have not changed appreciably during the last two decades. The shortcomings of these technologies become more and more clear and new architectures of patient monitoring devices are needed. This is discussed in the third section.

Furthermore trends in healthcare such as ageing populations, exponentially rising costs of care and shortage of experienced medical staff require changes in the conservative medical field. Advances in hardware, IT and data analysis technologies are enormous. New technologies may come to help to support clinicians. The main technologies, their status, the trends and outlook for the near future of patient monitoring are discussed in the last section.

10.1 Time Critical Care and High-Acuity Monitoring in the ICU

The intensive care unit (ICU) is a department of the hospital that is specialized in the treatment of patients with high risk conditions. In these critically ill patients, vital functions are threatened and time critical care is needed for survival. Patients enter the ICU after triage in other departments of the hospital such as the emergency department, surgical department or the ward. In the ICU the level of care is highest and it has the most advanced technological instrumentation (see Figure 10-1). Patient monitoring is only a part of the technical instrumentation in an ICU room. Please note the large number of therapeutic devices needed to support lung ventilation and administration of fluids, food, gases and medicines.

Patients are treated by highly specialized physicians (intensivists) and nurses (intensive care nurses). The nurse/physician to patient ratio is close to a one-to-one ratio. When needed doctors and nurses are available in seconds. In the ICU the highest level of continuous monitoring is needed and blood samples are taken at regular time intervals. Continuous information of blood pressure(s), oxygenation (both arterial and venous) and cardiac output is often required. Therefore invasive measurements are common, in the ICU the benefit of highly invasive measurements clearly

outweighs the risk. Furthermore patients are sedated or patients are too weak to breathe by themselves and mandatory or assisted mechanical ventilation is common. This requires monitoring of the airway pressure, flow, tidal volume and capnography. Often patients are sedated, the level of sedation needs to be adapted and sedation medicines can have a large (negative) impact on hemodynamics. Furthermore medicines, nutrients and fluids are injected into the large veins and arteries by infusion systems. In previous chapters monitor systems in the ICU, central monitoring stations, alarms issues and mechanical ventilators have been discussed.

In this section two examples of ICU care, therapies, monitoring and use of data are discussed. These are the most common and challenging disease conditions in the ICU. The focus in this section is on hemodynamic shock states and the acute respiratory distress state (ARDS, COVID-19).

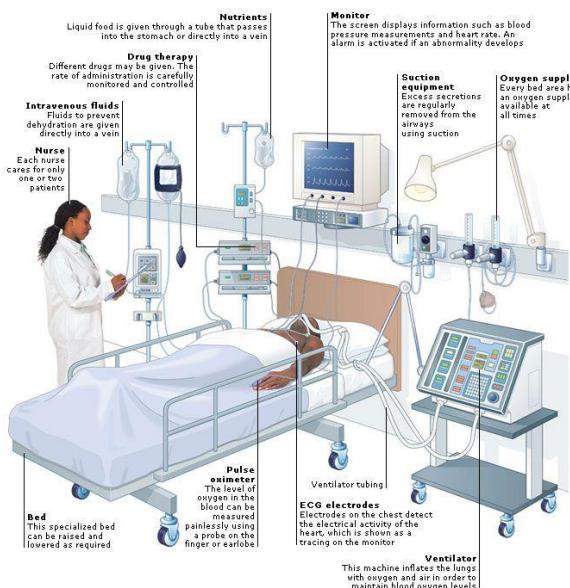


Figure 10-1 Schematic drawing of patient care in an ICU unit.

10.1.1 Shock

Shock is a medical emergency condition where tissues and organs are not receiving a sufficient flow of blood. If this shock condition lasts too long, severe and potentially permanent organ damage can occur and the probability of permanent disability or even death is large. There are different types of shock each with a specific cause. The main types of shock are illustrated using Figure 10-2.

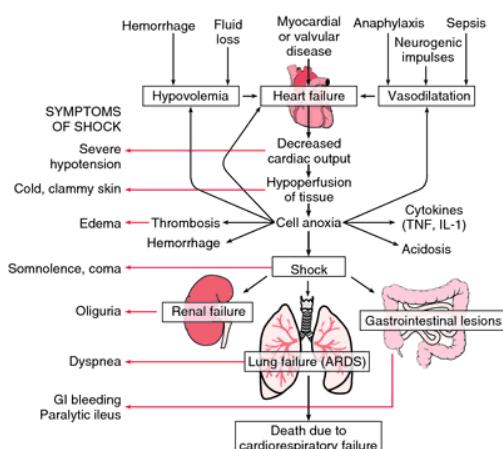


Figure 10-2 Diagram to illustrate the various types of shocks, the symptoms and possible consequences.

In the hypovolemic type there is a shortage in blood volume which leads to a decrease in cardiac output and finally blood pressure. In a hemorrhagic shock there is a severe loss of circulating blood volume due to for instance injury, trauma or surgery. This causes hypovolemia, a reduced stressed blood volume and as a result a strong decrease in blood pressure, cardiac output and finally in an insufficient perfusion of organs. This causes organ failure, this is a life threatening condition. Fluid loss due to continuous vomiting and diarrhea is another type of hypovolemic type of shock.

In the second type of a cardiogenic shock, the heart pump function is degraded (myocardial infarction, heart failure, valvular disease), both cardiac output and blood pressure decrease, and injury of the kidneys is a common complication. The final stage may be a lethal positive feedback spiral towards lower and lower cardiac output and subsequently death.

The most common type of shock is the distributive type. A severe vasodilation in large and small arteries and arterioles and veins can be triggered by a plurality of processes. Vasodilation can for instance be caused by an overreaction of the immune system to infections, overstimulation of the parasympathetic central nervous system, by anesthesia agents and by allergies (insect bites, food). In a shock state all organs require maximum perfusion. The sum of these flows exceeds the maximum cardiac output of the heart and as a result blood pressure decreases to levels below the value needed for organ perfusion, this results in hypoperfusion of organs. Arterial blood pressures should be larger than the minimum level required for organ perfusion (systolic pressure ~60 mmHg). When this hypotension period lasts too long multiple organ failure occurs. The probability of survival depends on the level of organ damage and the number of affected organs. When three or more organs are affected the mortality rate is very high (> 50%).

The most important form of distributive shock is sepsis, a life-threatening medical condition caused by the over response of the body immune system to an infection. When the concentration of the antibodies, hormones and chemicals released by the immune system is too large a severe form of vasodilation can occur, blood pressure will drop to too low levels and the body cannot compensate anymore. Subsequently a multiple organ hypoperfusion will occur and organ failure can lead to a high mortality rate. It is important to note that for Sepsis an increase in cardiac output precedes the decrease in blood pressure.

Finally there is a less common type of distributive shock, obstruction of blood flow can occur for instance due to cardiac tamponade, pneumothorax or pulmonary embolism.

The time in which a shock state can develop and progress to a very severe disease state is short (hours to days), the right therapy needs to be administered as early as possible. Early detection of a shock state is of vital importance. *This is still a major issue and one of the most urgent clinical needs.* There is a large research activity to detect hemodynamic deterioration, there is still no good solution. Note that different types of shock (cardiac versus distributive) need a very different therapy, administration of the wrong type of medicines is very dangerous and potentially lethal. There are three time stages in a shock state and the impact and severity increases rapidly.

Stage I: A compensated, non-progressive stage. The body systems act to maintain blood pressure. The body reactions are: faster heart rate, increase or decrease in vascular tone, vasoconstriction or vasodilation, kidney retains body fluids (low urine output), cardiac output changes, and faster breathing rate in reaction to hypercapnia. Timely treatment can halt further progression.

Stage II: A decompensated and progressive state. Body systems have reached limits, there is a too low perfusion, resulting tissue hypoperfusion, hypoxia, hypercarbia, lactate levels increase (anaerobic stage), patient becomes confused, consciousness level decreases, organ function

decreases (heart, lung, kidneys), blood coagulation occurs, there is a reduced micro circulation, damage to micro circulation, severe local hypoperfusion, positive feedback cycle, things get worse (rapidly) but quick and adequate treatment may halt progression.

Stage III: irreversible changes have happened, length of shock and hypoperfusion were too long. Things get worse rapidly, cardiac output decreases further, permanent damage to cells followed by cell death, organ failure, multiple organ failure and finally death.

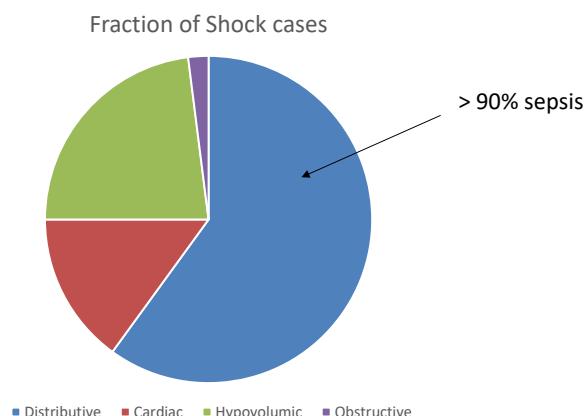


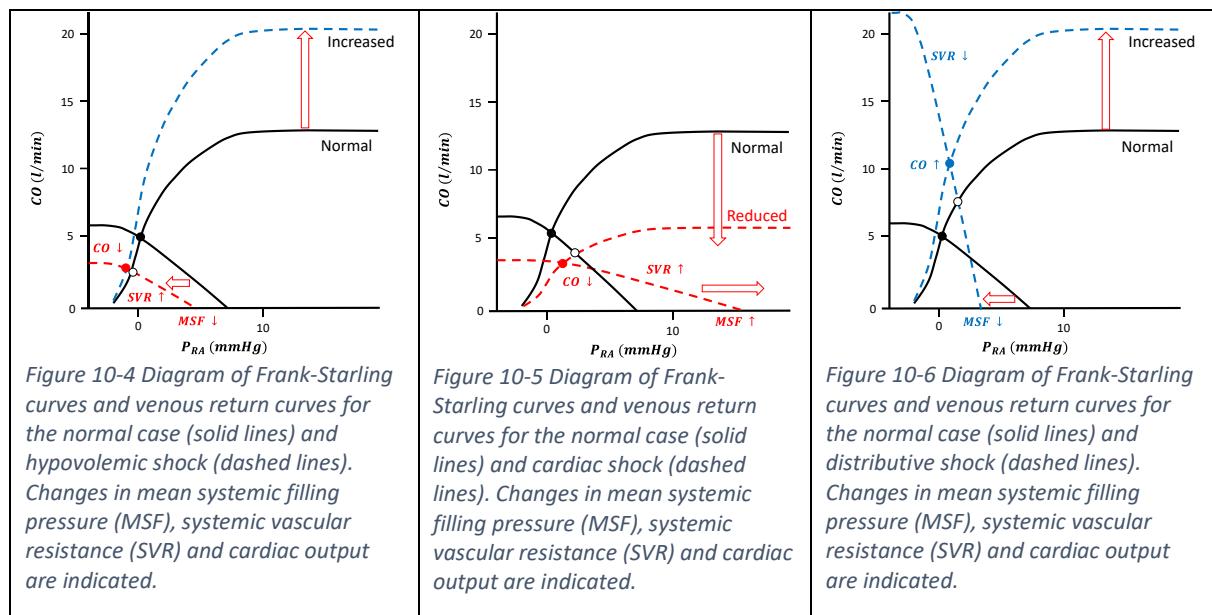
Figure 10-3 Diagram showing the distribution in numbers of the various shock states.

Note again that for all forms of shock early detection is of vital importance and early monitoring and diagnosis and subsequent treatment are very important. The prevalence of the different types of shock is shown in Figure 10-3. The distributive type of shock related to vasodilation is most common (> 60% of all types of shock) and in this category sepsis is the most common (90% of the cases). The hypovolemic type of shock (>20%) and cardiac shock (< 15%) also occur frequently.

10.1.2 Therapy during shock states

The optimal therapy for the different types of shock is illustrated using Guyton model curves for venous return and stroke volume. In Figure 10-4, Figure 10-5 and Figure 10-6 three schematic diagrams of the Guyton model curves for venous return and cardiac output (Frank-Starling curves) are plotted for the three dominant shock types.

For the hypovolemic shock type (see Figure 10-4) blood or fluid loss from the vascular system caused a decrease in stressed blood volume and a reduction in mean systemic filling pressure (MSF). This causes a reduction in cardiac output and the body reacts with strong sympathetic stimulation by the central nervous system. This causes increased heart rate, increased contractility of the heart, increased vascular tone and centralization of blood flow, i.e. switching off blood flow to the periphery and redirecting blood flow to the brain, heart and kidneys. This increases the systemic vascular resistance (SVR). The dashed lines in Figure 10-4 indicate the changes in venous return and the cardiac output curves. The cardiac output decreases and despite a maximum effort of the heart only little improvement in the cardiac output is observed. The therapy for this type of shock is based on restoring the unstressed blood volume by blood transfusions and IV fluid administration. The use of vasopressors is not recommended as there is already maximum sympathetic stimulation. The heart function is good and no measures are needed to strengthen the heart.



A cardiac shock is a result from existing heart failure or heart failure induced by a myocardial infarction. The pumping function of the heart is strongly reduced. In Figure 10-5 the venous return and Frank-Starling curve before and after the reduction in pumping function are shown. Directly after the event the cardiac output is strongly reduced (open symbol) and maximum sympathetic stimulation occurs. However the heart pumping function is strongly degraded and the venous return cannot be pumped by the heart, the right atrial pressure and the mean systemic filling pressure increase. Furthermore due to blood flow centralization the systemic vascular resistance is very high. The therapy for cardiac-shock is aimed at restoring the heart pumping function by administration of inotropes (medicines that increase the strength of the cardiac contraction) and by revascularization of blocked coronary arteries. Adding fluids or vasopressors would worsen the cardiac shock, lead to mortality and they should not be administered during therapy of the cardiac shock. In some cases an aortic balloon pump is inserted in the thoracic aorta to reduce the workload of the heart and buy time for further treatment.

For the distributive type of shock (see Figure 10-6) there is a severe vasodilation. This causes a reduction in the mean systemic filling pressure as stressed blood volume decreases due to an increase in lumen diameter and reduction in vascular compliance. The release of vasodilatory compounds causes a strong reduction in systemic vascular resistance. This causes a large shift and leftward rotation of the venous return curve. The cardiac output increases but not enough to maintain a sufficient blood pressure. The heart rate will go up and contraction strength increases, and cardiac output increases, but not enough to increase the blood pressure to acceptable levels. The therapy is aimed at increasing the mean systemic filling pressure by IV fluid administration and by adding vasopressors to increase the systemic vascular resistance and increase the vascular tone (decrease venous and arterial lumen diameter and compliance).

A summary of the therapies for hemodynamic stabilization during the different shock states is shown in Table 3. The changes in parameters and therapeutic agents are summarized also in Table 3.

	Right Atrial Pressure	Cardiac output	SVR	Therapy
Hypovolemic shock	↓	↓	↑	IV fluids
Cardiogenic shock	↑	↓	↑	Inotropes Revascularization
Distributive shock	↓	↑	↓	Vasopressors IV fluids

Table 3 Table showing the main changes in hemodynamic parameters of the Guyton model. The red arrow is the primary deviation from normality. SVR is the systemic vascular resistance.

The short term time critical care treatment is aimed at stabilization of the cardiovascular and respiratory systems to buy time for further treatment. Aggressive antibiotics are given to sepsis patients with a bacterial infection. Blood and fluid loss should be limited for hypovolemic shock. The cardiac pumping function of the heart must be improved for the cardiac shock case. The monitoring requirements during treatment of shock in the ICU are discussed in the next section.

10.1.3 Monitoring and Measurements during Shock States in the ICU

The first aim during shock treatment is to maintain hemodynamic stability and oxygen delivery to the tissues and organs to buy time for further treatment and to limit further tissue and organ damage. In Table 4 the main parameters, acceptable ranges of parameters and its measurements are shortly recapitulated. The number one priority is sufficient oxygen delivery to tissues and organ perfusion. Oxygen delivery requires cardiac output measurements, preferably continuous cardiac output and measurement of arterial blood composition (hemoglobin concentration and its saturation with oxygen). The arterial blood composition requires blood samples. Continuous non-invasive SpO₂ is a good surrogate for SaO₂. Note that a central site oximetry sensor is preferred during shock states. The mean arterial pressure (MAP) is an important parameter for organ perfusion. The autoregulation of blood flow in organs requires a MAP greater than 60 mmHg, for smaller MAP values blood flow through vulnerable organs (Kidney, brain...) is reduced. A MAP value greater than 65 mmHg is recommended. A reliable measurement of MAP requires a continuous invasive arterial probe. The relation of cardiac output and central venous pressure (CVP) in the Guyton model has made CVP an important parameter during shock treatment. However many studies have shown that in clinical use cases the CVP parameter has little relevance and its use has been questioned. The pulmonary capillary wedge pressure (PWCP) is a surrogate of the left atrial pressure, it is a useful parameter only when pulmonary hypertension and lung edema is suspected. Note that a Swan-Ganz catheter can be used for cardiac output, CVP and PCWP. However the use of this catheter poses high risk to the patient and its use has strongly reduced. The last three parameters are related to ventilation, oxygenation, oxygen consumption and blood composition. The tidal volume, driving pressure and plateau pressure are important parameters for lung protective ventilation. They are monitored most often on the ventilation machine but measurements are often copied to the patient monitor.

Simultaneous waves of the ventilation and circulation parameters provides useful information to the clinician. The last two measurements are related to oxygenation, oxygen uptake by the tissues and acidosis of blood, the measurements require blood sampling from the central venous site and a large artery. For acidosis the pH and partial pressure of CO₂ and concentration of bicarbonate ions are important. The pH is important for the functioning of many processes and proteins and should be monitored. The ETCO₂ parameter that is measured at the airway opening is a surrogate for arterial CO₂ partial pressure and also provides indirect information on metabolism and blood flow. The arterial oxygen concentration and saturation are important for oxygen delivery, the saturation of venous blood provides important information on tissue oxygen consumption and capillary blood

flow. During shock the micro circulation may be impaired and arterial blood may flow for a large part by shunts vessels to the veins. In this case the oxygen concentration in the veins is much higher than the normal condition and measures need to be taken. Lactate is a marker of insufficient oxygen supply to the cells and tissues, it is a sign of tissue hypoxia. This is an important marker as microcirculation can be disturbed and oxygen saturation in the large blood vessels can still have normal values. It is also measured in blood samples. Note that lab measurement of many types of biomarkers and other substances from blood samples provides important information to the clinician that is essential for treatment optimization.

	Parameter	equation	Value
Oxygen delivery	DO_2	$DO_2 = S_a O_2 \cdot [Hb] \cdot CO$	> 250 ml/min
Organ perfusion	$MAP, \text{urine output}$	$MAP = CO \cdot SVR$	> 65 mmHg > 500ml/day
Central venous pressure	CVP	CVP waveform	8 – 12 mmHg?
Pulmonary capillary wedge pressure	PCWP	Pressure waveform	10 – 14 mmHg
Ventilation	Tidal volume, Plateau pressure	Pressure waveform	6 ml/kg PBW < 30 cmH ₂ O
Acidosis	pH, pCO ₂	$pH = 6.1 + \log \frac{HCO_3^-}{0.03P_{aco_2}}$	7.35-7.45
Oxygenation	pO ₂ , $S_a O_2$, $S_{cv} O_2$, lactate	Arterial and venous partial O ₂ pressure, saturations, lactate concentrations	$S_a O_2 > 90\%$ $S_{cv} O_2 > 70\%$

Table 4 Main parameters, equations and target values measured during shock states.

Fluid administration is an important part of a therapy to maintain hemodynamic stability. Optimization of fluid therapy is one of the main challenges during time critical care. It is discussed in more detail in the following section.

10.1.4 Fluid Therapy

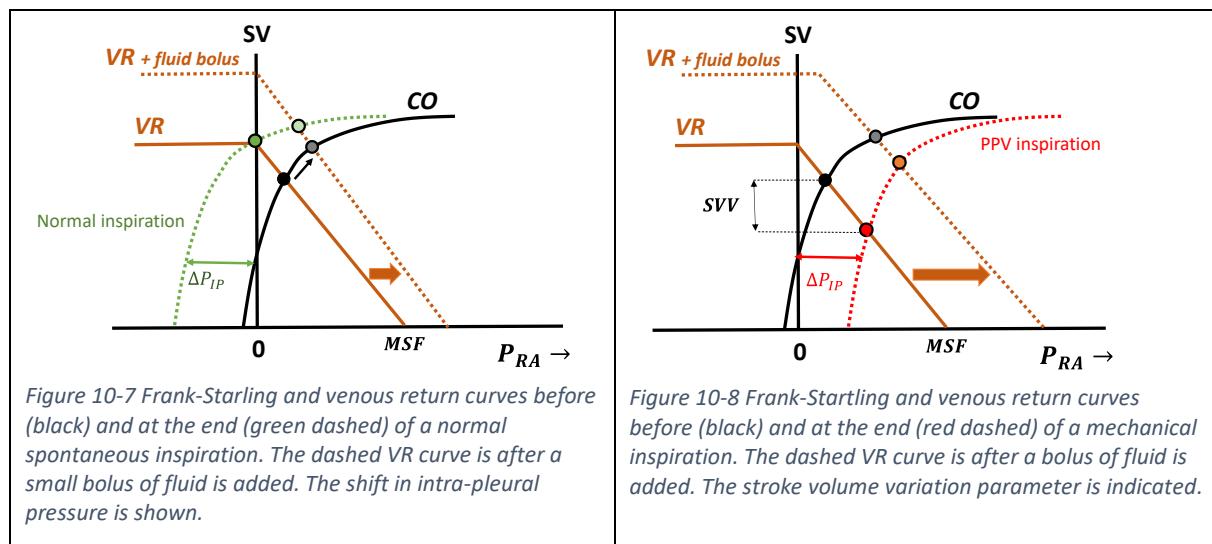
Fluid therapy is used in shock states but also during surgery, trauma, fluid loss and several other causes. Most often blood pressure is low and this is an indication that fluid therapy is needed. The fluids are used as volume expanders. The use is widespread, there are however many issues as will be discussed below. Too little or too much fluid is both harmful. The two most used types of IV fluids are crystalloids and colloids (see Table 5).

Most fluids contain appreciable concentrations of salts, sodium chloride is common. Note that the recommended daily intake of sodium chloride is 6 grams for a healthy adult. Note that 1 liter of 0.9% saline solution corresponds with 150% of the daily intake. In some cases more than 10 liter of IV fluid per day is used. In larger amounts this type of fluid is harmful, especially for patients with critical illness with already impaired kidney function. Excessive fluid administration has been linked with acute kidney injury and increased mortality. This also holds for the other fluids, all have side effects and many studies have shown they are toxic when large amounts of IV fluid are administered. Too little fluid does not lead sufficient volume expansion and increase in blood flow and arterial pressure. Too much is detrimental and excessive use has been linked with increased mortality. This is an important issue that is still studied by many researchers. There is a need for monitoring techniques, methods and support techniques that assist the clinician in decision making during fluid therapy. There is no optimal solution yet.

Types		Composition
Crystalloids	Saline solution	$NaCl 0.9 \% / l$
	Ringer Lactate	$NaLct 3.1 \frac{gr}{l} NaCl 6 \frac{gr}{l} KCl 0.3 gr/l CaCl2 0.2 gr/l$
	Glucose, Dextrose	50 gr/L + salts
Colloids	Gelatin	35 – 60 gr/L + salts
	Starch	40 – 80 gr/L + salts
	Albumin	50 gr/L
Blood	-	Whole blood
Blood	components	Plasma, RBC, WBC, platelets

Table 5 Common IV fluids and composition(s)

There are two important questions when fluid therapy is applied, how do we know that IV fluids are needed (is the patient fluid responsive), is therapy effective and secondly how do we know that the right amount of fluid has been administered. As mentioned before there is no optimal method available, the most used tests are described below.



Note that initially the central venous pressure was used as the main indicator of hemodynamic stability but many studies have shown that this parameter was of little value and one review study concluded that measuring CVP to determine if fluid therapy was needed was not better than flipping a coin. The Guyton model is used to illustrate the three most used methods. The most common fluid responsiveness test is to add a small bolus of fluid (250 ml to 500 ml) in to the circulation and monitor the response in cardiac output or stroke volume (see Figure 10-7 and Figure 10-9). Note that a continuous cardiac output measurement is preferred to guide therapy. When this is not available arterial blood pressure is used as a surrogate.

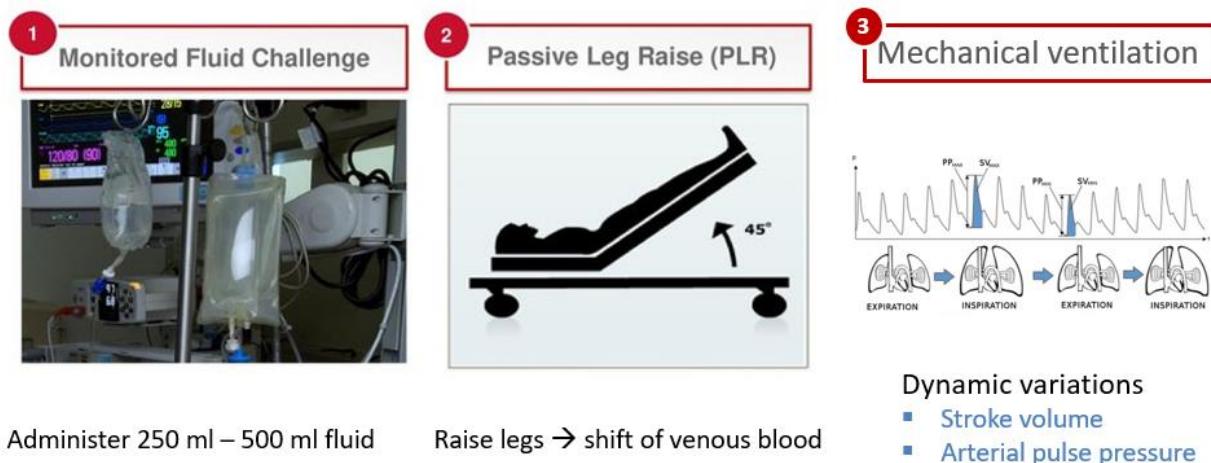


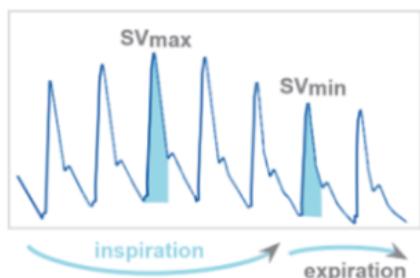
Figure 10-9 Most common used methods to detect fluid responsiveness.

When an increase in stroke volume (typically 15%) and/or arterial pressure is observed within 15 to 30 minutes after the bolus administration, the clinician can decide to add another bolus of IV fluid and can continue until desired values of vital sign parameters are reached. An issue is that most healthy persons are fluid responsive when this method is applied, fluid responsive does not mean that IV fluids must be added. Administration of vasopressors could have been sufficient.

In the passive leg raising technique blood will flow from the large veins in the legs to the abdomen and thorax and extra blood volume flows in to the right heart. The excess blood volume will be pumped into the arterial system. Observing an increase in stroke volume, cardiac output or arterial blood pressure can then be used by the clinician to decide to add extra IV fluids. Criteria are similar to those of the monitored fluid challenge.

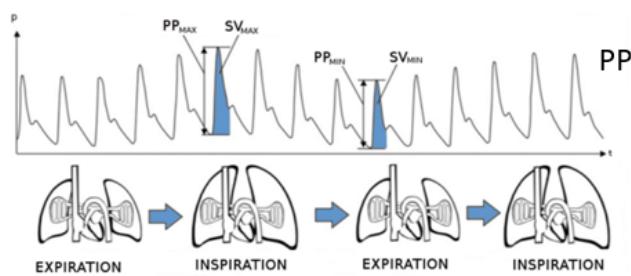
The third method to guide fluid administration uses dynamic fluctuations in arterial pressure or stroke volume in mechanically ventilated patients (see Figure 10-8). The increase in intrapleural pressure during inspiration will lead to two effects. First the rise in intrathoracic pressure will lead to an immediate increase in left ventricular stroke volume. Secondly the right shift of the Frank-Sterling curve leads to a lower filling of the right ventricle and to a lower RV stroke volume. This decrease in stroke volume will appear after a few heart beats in the left ventricle and to a reduced LV stroke volume during the expiration phase of the mechanical ventilation. These changes in LV and RV stroke volume during positive pressure ventilation lead to a stronger modulation of the LV stroke volume and arterial pressure (see Figure 10-10). From the modulation the pulse pressure variation parameter (PPV) and stroke volume variation (SVV) can be determined. Whenever these parameters exceed a threshold value (PPV, SVV > 14% - 19%) this is an indication that the patient may be fluid responsive. Adding a bolus to the circulation can be considered and changes in the dynamic SVV and PVV parameters can guide the fluid resuscitation therapy. The drawback of the technique is that mechanical ventilation with a relatively large tidal volume is needed and in many cases this is not possible or desirable because of the negative impact on cardiac output or increased injury of the lung due to the larger tidal volume and pressures. Furthermore studies have shown that the specificity and sensitivity are not as good as the first two methods.

Stroke Volume Variation (SVV)



$$SVV = \frac{SV_{max} - SV_{min}}{SV_{mean}}$$

Pulse Pressure Variation (PPV)



$$PPV = \frac{PP_{max} - PP_{min}}{PP_{mean}}$$

Figure 10-10 Stroke volume and pulse pressure variations during positive pressure mechanical ventilation.

During and after hemodynamic stabilization, the therapy is aimed at removing the root cause of the shock. In case of bacterial sepsis broad spectrum anti-biotics are administered as soon as the sepsis diagnosis is made. Blood cultures tests are performed to identify the type of invader that caused sepsis. Thereafter medicines can selected to optimize the elimination of the foreign pathogens. Bacteria, yeasts and other microorganisms need to be identified as soon as possible. This may take days and there is a high need for short term identification using biomarkers blood samples only. The therapy is adjusted when the severity of the disease state changes. Single monitor parameters do not give sufficient information on the disease state. Scoring systems with multiple parameters and with certain biomarkers have been developed and are widely used in clinical practice.

10.1.5 Severity of disease Scoring systems in the ICU

Single monitor vital sign parameters do not provide sufficient information on the health status of a patient. Furthermore patient monitor data lacks information on organ function as can be obtained from biomarkers obtained from blood samples. The estimate the severity of a disease state multi-parameter scoring systems have been proposed. The total score indicates the severity of the disease and is related to the mortality. The two most used systems are discussed below.

APACHE II ("Acute Physiology And Chronic Health Evaluation II") and SOFA ("Sequential Organ Failure Assessment") are scoring systems used in the ICU to estimate and quantify the severity of the disease state, to estimate probability of mortality and guide therapy. They are based on large clinical studies and statistical analysis of data obtained in these studies. The Apache II scoring system was first published in 1985. A table with physiologic variables and scoring per actual value is shown in Figure 10-11.

The data is gathered within 24 hour of the admission in the ICU. The list included the main vital sign parameters, blood gas parameters, Sodium and Potassium ion concentrations in serum, Creatinine in serum as a measure for acute renal injury and white blood cell count as a measure for the function of the immune system. The Glasgow coma scale is used to estimate the level of consciousness, note that this scale can have a large impact on the Apache II score. It is a very important parameter. Furthermore age and chronic health condition are included in the total score. Furthermore, as expected age is a very important parameter. In the predicted death rate as function of the Apache II score is shown. The death rate increases strongly with the score.

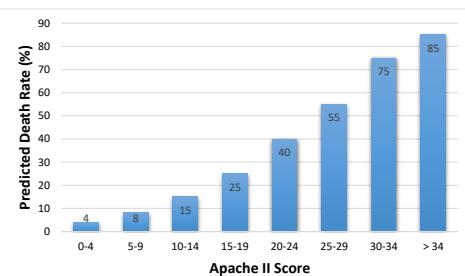
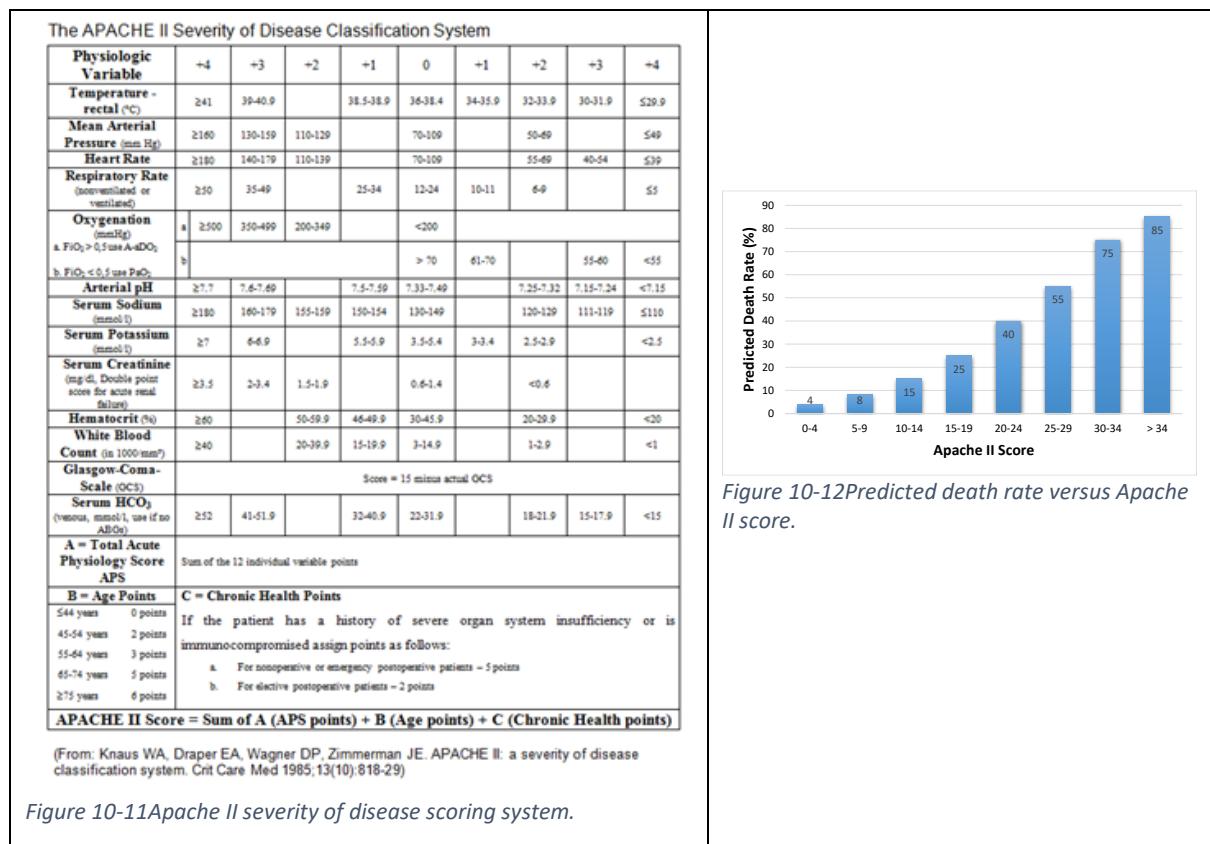


Figure 10-12 Predicted death rate versus Apache II score.

The SOFA score (“Sequential Organ Failure Assessment”) previously known as “Sepsis related Organ Failure Assessment” score was developed to study the extent to which organ function or organ failure has developed. The score is based on the functioning of six different organ systems, the cardiovascular system, the respiratory system, hepatic (liver), renal (kidney), blood (coagulation) and neurological systems. The scoring system is shown in Figure 10-13. The predicted hospital mortality is shown in Figure 10-14.

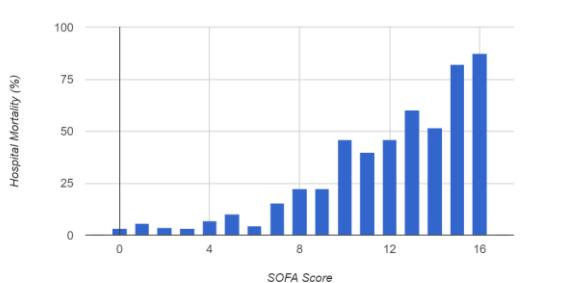
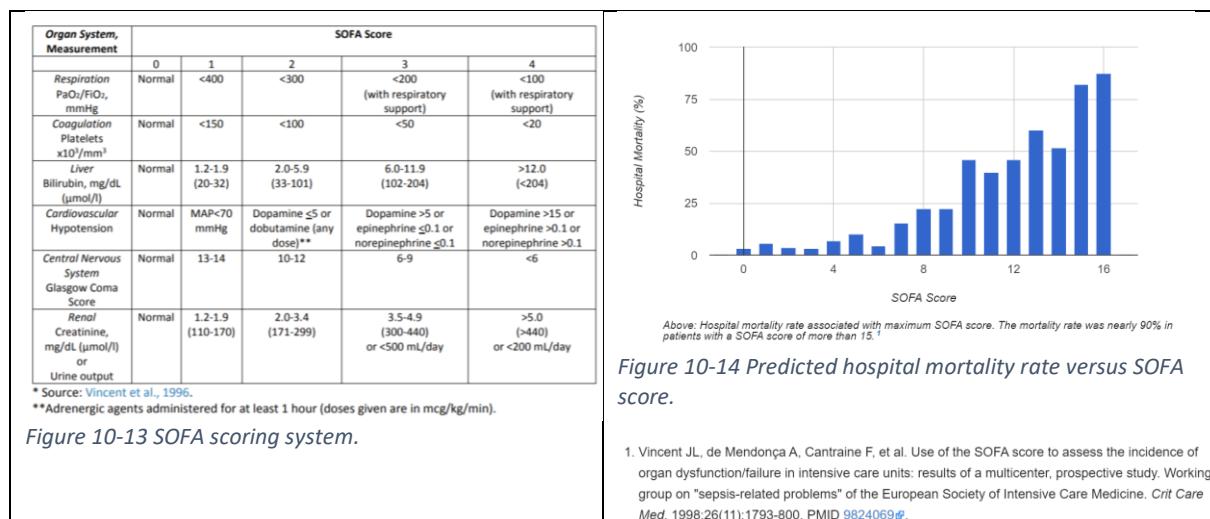


Figure 10-14 Predicted hospital mortality rate versus SOFA score.

1. Vincent JL, de Mendonça A, Cantraine F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on “sepsis-related problems” of the European Society of Intensive Care Medicine. Crit Care Med. 1998;26(11):1793-800. PMID: 9824069.

Both scoring systems require many measurements, blood sampling is required. These test are complex and time consuming. They are restricted to the ICU. Furthermore these tests are based on ICU population averages and probability of mortality cannot be predicted for a specific patient.

There is a need for a fast screening technique or techniques with a better sensitivity and specificity. The Quick SOFA (qSOFA) was recently proposed to early identify patients in the ICU with a high risk of poor outcomes with an infection. It is shown in Figure 10-15. The test includes only respiratory rate, low systolic blood pressure and change in mental status (measured using the Glasgow Coma Scale).

qSOFA (Quick SOFA) Criteria	Points
Respiratory rate $\geq 22/\text{min}$	1
Change in mental status	1
Systolic blood pressure $\leq 100 \text{ mmHg}$	1

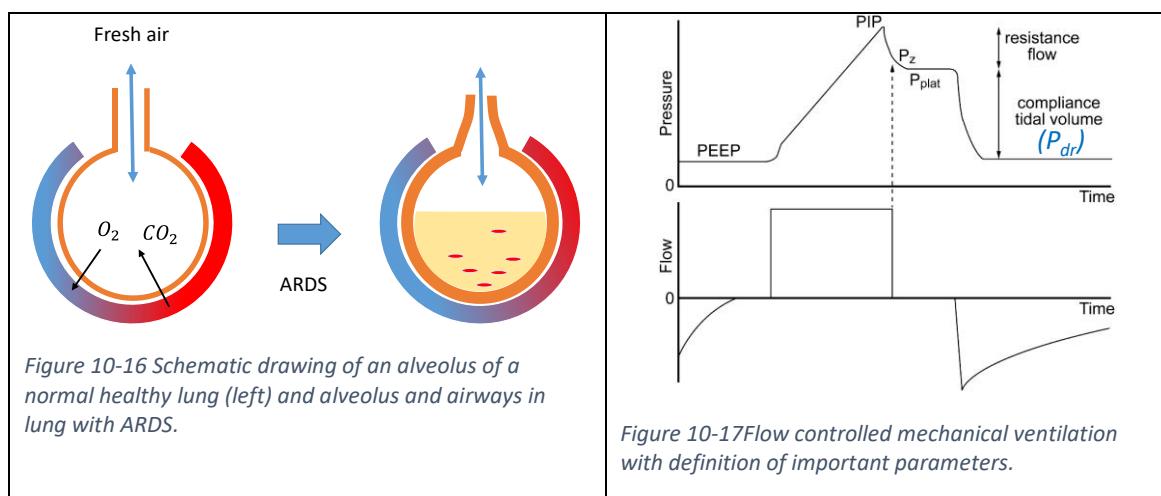
Figure 10-15 Quick SOFA scoring system.

A qSOFA score of 2 points correlates with a high probability of mortality or increased length of stay in the ICU. The qSOFA technique is recommended for use outside the ICU to identify patients at risk for sepsis. Note that these scoring systems still need further improvements, and there is a need for improved scoring systems that are more patient specific. It is well accepted that early recognition and timely treatment largely determine outcome of sepsis.

During sepsis lung function may be impaired and acute respiratory distress syndrome may develop. A similar condition can develop during COVID related lung infections. ARDS and COVID-ARDS are briefly discussed in the next section.

10.1.6 Acute Respiratory Distress Syndrome

The acute respiratory distress syndrome (ARDS) is a non-local and acute inflammation of the lungs. The first signs are shortness of breath, a rapid respiration rate and a reduced arterial oxygen saturation. ARDS impairs oxygen and CO₂ exchange and affects lung tissue as well as capillaries. The mortality rate is high, 30 to 60% of the patients die. The survivors have often impaired quality of life. As for sepsis early detection is of vital importance.



ARDS is a complication of a primary disease, a reaction of the lungs to a severe underlying disease such as pneumonia, COVID, sepsis, aspiration, mechanical ventilation, trauma and others. It causes a violent inflammatory reaction in lungs for reasons that are poorly understood. Fluid, proteins and inflammatory cells from the pulmonary blood vessels enter the lung tissue. Blood flow through the capillaries is affected, pulmonary blood flow resistance may increase. As a result of both local lung

inflammation and restriction in blood flow the ventilation-perfusion ratio is inhomogeneous. The results of progressed ARDS are: fluid filled stiff lung(s), thicker alveolar walls, increased airway resistance and local collapse of airways (atelectasis) (see Figure 10-16). It is difficult for air to reach the alveolar space due to higher airway resistance, the stiff lung limits acceptable tidal volumes unless high airway pressures are applied. The impaired gas exchange requires a PPV mechanical ventilation, with a sufficiently high PEEP pressure to open the airways, recruit aerated lung volume and alveoli and to reduce airway resistance. The inspiratory gas has a high inspiratory fraction of oxygen (FiO_2) to improve oxygenation of pulmonary blood. The lungs are very susceptible for barotrauma and volutrauma. Protective mechanical ventilation (low tidal volume, low pressures)) is needed to maintain oxygenation of arterial blood and oxygen delivery to tissues and to buy time for recovery. The following ventilation parameters and maximum values are recommended for lung protective ventilation during ARDS.

1. Tidal volume 6 ml/kg predicted body weight (PBW \leftrightarrow length)
2. Plateau pressure < 30 cmH₂O (pressure when flow is absent)
3. Limit driving pressure $P_{dr} < 14 \text{ cmH}_2\text{O}$
4. Higher PEEP > 5 cmH₂O (keep airway open, limit atelectasis)
5. Prone position
6. Muscle relaxants (initially)

The remaining therapy is aimed at stabilization, treating the root cause and to buy time for recovery. Often the inflammation is caused by an overreaction of the immune system and this overreaction finally kills the patient. For virus infections medicines can be given to reduce the overreaction of the immune system, there is no other therapy (i.e. COVID). Sometimes formation of blood clots must be suppressed to avoid lung embolism (i.e. COVID). In other cases there is time for a therapy to remove the cause of ARDS. Note that often a bacterial or virus infection is the cause. Administration of broad spectrum antibiotics is common for bacterial infections. Blood cultures are tested to identify the specific bacteria and fine tune antibiotic therapy.

As noted before there is a need for early detection of Sepsis and ARDS in lower acuity settings like the ward. Note that a large fraction of the patients admitted in the ICU comes from the hospital ward. The number of cardiac and respiratory arrests is largest. In the ward there is a very low level of monitoring of patients and this increases the risk of severe illness being detected too late or not at all. Monitoring in the ward could provide data that can be used in early warning systems and increase patient safety. The state of the art of this field is discussed in the next section.

10.2 Monitoring in the ward

Mortality in the perioperative tract accounts for a large fraction of total mortality but there are more factors that need to be taken into account. Studies in the late 1980's and 1990's have shown high rates of unanticipated complications, unplanned ICU admission, unexpected death and cardiac arrests occur in the wards of hospitals. The main findings of these are summarized in [Figure 10-18](#) and in the quote from Daryl Jones below.

*"Studies from the 1990s to the present day show that adverse events including cardiac arrest, unplanned admission to intensive care, and unexpected death in hospitals around the world are usually preceded by objective signs of deterioration, often for several hours" (Daryl Jones et al., *Intensive Care Med* (2016) 42:593–595)".*

It appears that the high number of patients deteriorating in the ward is an important factor for the overall hospital mortality. Deterioration is often unnoticed and can lead to life-threatening complications and even death. The studies also showed these life-threatening conditions are often preceded by observable signs of deterioration, often for many hours in advance of the event.

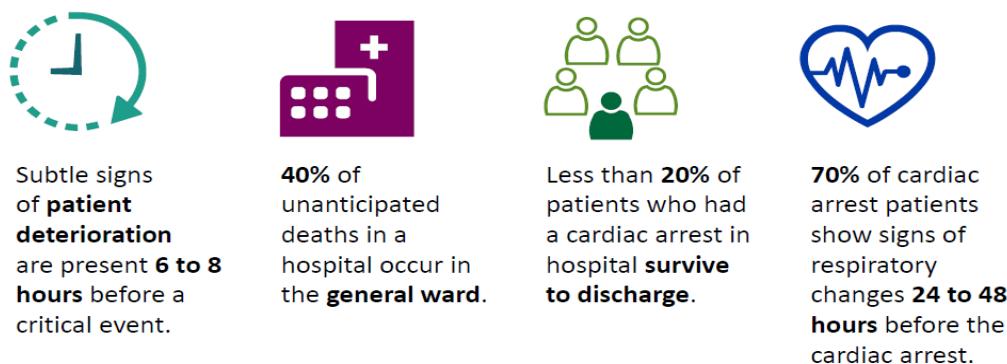


Figure 10-18 Summary of findings of studies of patient deterioration in the ward.

There are multiple causes for this high rate of unexpected complications and deaths but an important factor is the limited vital sign monitoring in the ward and the low training level of ward nurses and doctors on critical care and interpretation of monitoring results. ***Early detection of the deteriorating patient and imminent intervention could have prevented a cardiac arrest or reduced the impact of the complication like for instance sepsis. Time to treatment is crucial for the outcome.***

It should be noted that at present patients in the ward are older, sicker, have several comorbidities and have a higher risk for deterioration. Furthermore these patients are more susceptible for hospital acquired infections that can lead to sepsis and other severe diseases. Note that the work load of the nurses in the ward is already too large, there is little time for extra tasks. This high workload leads to errors in for instance administration of drugs. Systems have been developed to cope with unnoticed deterioration of patients in the ward. As noted before early detection is key. These systems are discussed below.

10.2.1 Rapid Response Teams or medical Emergency Teams

Pioneering hospitals in Australia and the UK have recognized this problem in the ward and in the 1990's they proposed to solve this hospital wide problem by the introduction of so-called Rapid Response Teams (RRT) or Medical Emergency Teams (MET). The system is illustrated in the diagram shown in [Figure 10-19](#). It consists of four branches. Note that the terms afferent and efferent are defined in anatomy. As an example we use a nerve. A sensory nerve is also called an afferent nerve.

It directs stimuli to the central nervous system. The motor nerve or efferent nerve works the other way, relaying signals from the central nervous system. This jargon is used in the field of rapid response systems.

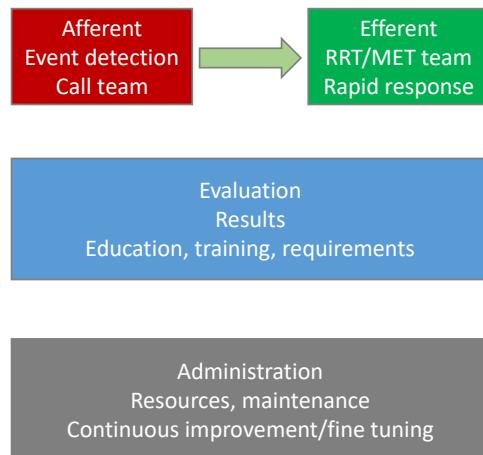


Figure 10-19 Diagram of a RRT/MET system. The four main parts are the detection (afferent part), the medical response team (efferent part), the clinical; system part and the organizational/administrative part.

The afferent limb is the part of the system that detects the event (sensory part). The afferent part consists of a nurse or ward staff or could be a (automatic) ward monitor system. The nurse activates a special team (RRT,MET) that will visit the patient within a few minutes (efferent limb, therapy). The team examines the patient, applies therapy and if needed decide to scale up the level of care. This team (the efferent part) consists of three people, typically an ICU, anesthetist or emergency department physician leads the team and is supported by critical care nurses. The team composition can vary but typically consists of clinicians with experience in critical care and respiratory care. Often the ICU or anesthesia department delivers the manpower for the RRT/MET team. The system further consists of two organizational units, a clinical part that continuously evaluates the results and fine tunes the criteria for calling to optimize the sensitivity and specificity of the system. This is very important because the successful implementation requires a long period of optimization and is hospital specific. Furthermore this part offers education and training activities. The fourth part is the management and administrative part. It supports the system financially, approves investments and takes care of the administration of data and results. Note that introduction of a RRT system requires substantial manpower and investments in monitoring hardware and IT systems. The weakest link of the system will determine its quality. Initially and possibly still today the afferent or detection part has been the weakest link. Trigger criteria for calling RRT/MET teams are described in the next section.

10.2.2 Criteria for RRT/MET calls – Vital Signs and other parameters

The most common conditions that trigger the rapid-response system are acute respiratory failure, acute cardiac failure, and acute changes in consciousness, hypotension, arrhythmias, pulmonary edema, and sepsis. Furthermore respiratory depression due to excessive use of pain medication is of importance. There have been many studies for trigger criteria for RRT/MET team calls. Most are based on simple vital sign parameters that have been obtained by analysis of spot check data and subsequent refinement and extensions of these studies. They are summarized in *Table 6*.

Respiration rate (earlier)	Heart rate (earlier)	Blood pressure (late)	SpO ₂ (later)
Non specific marker <ul style="list-style-type: none"> ▪ Respiratory depression ▪ Acidosis ▪ Cardiac origin ▪ Sepsis ▪ Kidneys 	Non specific marker <ul style="list-style-type: none"> ▪ Cardiac origin ▪ Sepsis ▪ Blood loss ▪ Low and high rates 	Non specific marker <ul style="list-style-type: none"> ▪ Cardiac origin ▪ Blood loss ▪ Sepsis ▪ Hypotension ! 	Non specific marker <ul style="list-style-type: none"> ▪ Cardiac origin ▪ Respiratory ▪ Limit 90% ? ▪ False alarms
Temperature (late) Non specific marker <ul style="list-style-type: none"> ▪ Sepsis ▪ Infections ▪ Not skin temperature! 	Urine output (complex) Specific marker <ul style="list-style-type: none"> ▪ Renal failure ▪ Acute kidney injury ▪ Difficult, invasive 	<u>Consciousness</u> (early, nurse) Important marker <ul style="list-style-type: none"> ▪ A: Alert ▪ C: Confusion ▪ V: Voice ▪ P: Pain ▪ U: Unresponsive 	<u>Nurse worry</u> (early, nurse) Important marker <ul style="list-style-type: none"> ▪ Patterns ▪ Behaviour ▪ Experience

Table 6 Vital signs measurements and methods that can be used in the ward to trigger a RRT/MET team.

Most studies find that a decreased or an increased respiration rate is an important indicator of patient deterioration. A low rate is an indication of respiratory depression. A high respiratory rate is an important marker of disease. It can be caused by many disease conditions, it is not specific for a disease type. An example could be acidosis caused by diseases of the respiratory or cardio-vascular systems. Other causes could be kidney failure , a severe infection (sepsis, ARDS, pneumonia) or blockage of a pulmonary blood vessel by a blood clot (lung embolism). The respiratory rate is a good indicator of forthcoming deterioration but as mentioned before it is not specific and is often of non-clinical origin such as measurement errors (motion artifacts) or due to anxiety or stress. Therefore this single parameter is not enough to trigger an RRT call. Other vital sign parameters need to be added to the monitoring system. It appears that heart rate, blood pressure and SpO₂ are useful. Again these parameters are non-specific for a disease type and could be related to many causes. The heart rate and respiration rate are earlier indicators for imminent deterioration than SpO₂ and blood pressure. For blood pressure hypotension most often correlates with RRT calls. In the later phase of cardiac failure, shock states and sepsis blood pressure drops, but this is a late call, often too late. This limits the value of this parameter. The same holds for temperature, whenever fever is detected it is mostly late in the disease process. It was found that urine output is a very good indicator of renal failure but it can only be measured when a Foley catheter is inserted in the urine tract. This is not a common procedure and is very uncomfortable for the patient. Two of the most important and underestimated parameters are the level of consciousness and nurse worry. These parameters are obtained from the nurse during the spot check or during other encounters with the patient. They are difficult to obtain and cannot be measured with automatic systems.

A single vital sign parameter cannot be used to trigger the RRT system, a large number of false alarms or warnings would be the result. Sometimes the limit of parameters are set at extreme values, this reduces the number of false positive warnings but also reduces the number of true alarms. To improve sensitivity and specificity, the data several vital sign parameters are combined in scoring systems. These systems are discussed in the next section.

10.2.3 Early Warning Scoring System

In the previous sections the need for monitoring on the ward and complexities in the triggering of a call for a RRT team were discussed. The early detection of a potentially dangerous deterioration of the health condition is a formidable challenge because of the high demands on sensitivity and specificity. The use of simple one-parameter static alarms (as for instance SpO₂ lower than 90%) would lead to a false positive alarm overload and the system will not function properly. A possible solution is a multi-parameter scoring system. In principle such a trigger system reduces the number of false triggering of the RRT system. Furthermore contrary to a cardiac or respiratory arrest the deterioration process in the ward is often slower and there is more time to react. Therefore a warning is generated instead of an alarm. This is more in line with the workflow in the hospital ward and reduces the overload on the hospital staff. A further advantage is that the ward staff can focus on patients with a high score, this reduces work load, more time is available for patients at risk and it likely to improve outcome. An acceptable delay to respond to a warning is in the range of a few minutes to 15 minutes. Many studies have been performed on EWS based trigger systems and several early warning scoring (EWS) systems have been introduced. The aim is to have both a sufficiently high sensitivity and specificity. A system that is used in many countries is a variant of the early warning system developed in the UK. The national EWS was introduced in the UK (NEWS). The scoring system is based on large studies with data from many hospitals and it was optimized for use on the general ward. In *Figure 10-20* a diagram of the NEWS scoring system and the clinical response that should follow the NEWS score is shown. It is designed for a spot monitoring system with an option to escalate the monitoring frequency when needed.

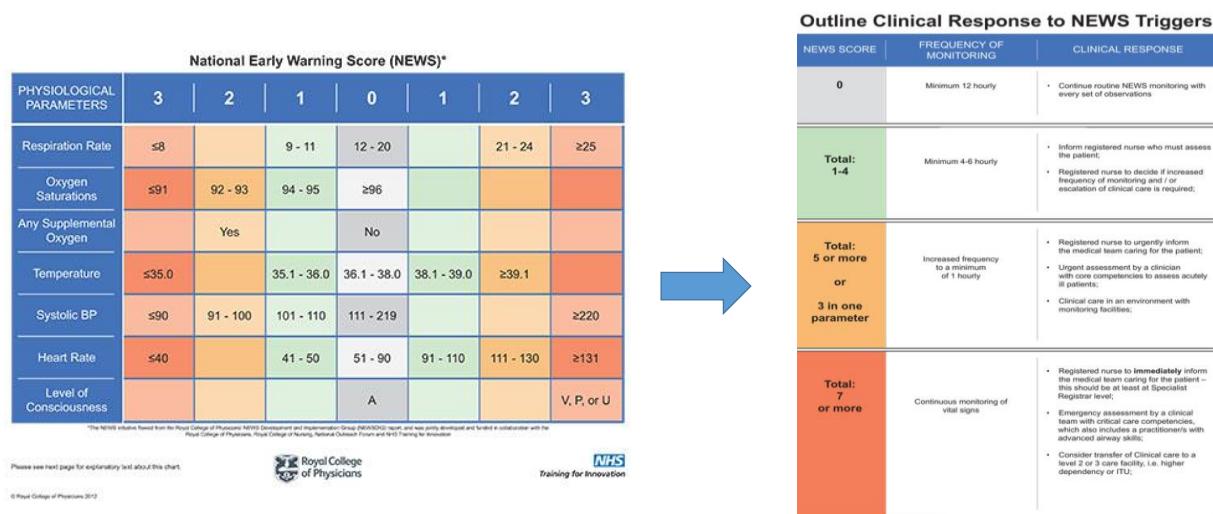


Figure 10-20 Diagram of the UK NEWS early warning scoring system for patients in the general ward.

Relevant physiological parameters included in NEWS are: the respiration rate, SpO₂ oxygen saturation, administration of supplemental oxygen, tympanic temperature, systolic blood pressure, heart rate and level of consciousness. For each parameter an ideal range of values is selected, it is the same for every patient. When the parameter is in the normal range the score is zero points. When the parameter is outside the normal range scoring points are given. The larger the deviation the higher the score. The maximum score is 3 point. The ranges and values of the scores are based on the results of large retrospective clinical studies and are chosen such that a good compromise between sensitivity and specificity is obtained.

The computed NEWS score is divided in four subcategories. For a NEWS score of 0 points no action is needed and the spot check monitoring frequency is reduced to two times per day. The next category

is a score between 1 and 4. A qualified nurse is notified and he/she must visit the patient. The nurse must decide if the monitoring frequency must be increased to once per four to 6 hours or if the level of care must be increased. For a score of 5 or more or when one of the parameters has a maximum score of 3 the qualified nurse must urgently inform the medical team which is responsible for the patient care, submit an urgent call to a physician with expertise in critical care and the patient may be transferred to an environment where more frequent monitoring is possible. When the NEWS score is higher than 7 the nurse should immediately inform the medical team at the specialist level and notice the RRT team. The team should include a critical care specialist and a specialist with advanced airway skills. Transfer to a higher level of care (medium care or intensive care) should be considered.

This system has an escalation in the level of urgency, level of monitoring and level of care. An advantage is that the scoring criteria can be adjusted by local clinicians and can be optimized for the specific hospital and characteristic patient groups. There are improvements needed. The first is the improvement of the quality of the data that is used to compute the NEWS score. The second is the registration of this score in the hospital information system. Manual data for the most critical parameters such as respiration rate are often less reliable and more advanced monitoring techniques are needed. However, monitoring in the ward differs from that in the ICU. Continuous monitoring is not needed and systems that are compatible with the general ward environment are needed. Furthermore the parameters in the system is not patient specific, this can be added if needed. Existing and new systems are discussed in the next section.

10.2.4 Monitoring Technology in the Ward

At present monitoring in the ward is mostly done by spot checks, i.e. two to three times daily by a nurse (see left picture of *Figure 10-21*). This is a time consuming task. Moreover the ward staff is not trained in the use of more advanced monitoring systems and in the use and interpretation of monitor data. The workload of the nurses is high, new techniques must reduce the workload, this is a difficult requirement. The investment budgets for monitoring hardware on the ward are limited, often management is not willing to invest. Finally patients in the ward are sicker and older than in the past, are often more or less mobile, have a high risk for falls. However patient are most often mobile, need to visit the toilet and take showers. The monitoring techniques for the ward must be compatible with mobile patients in a low-tech environment, with its lower training level of the staff and the workflow of the nurses. Finally they should be patient friendly and allow patient mobility. The following requirements are important:

- Monitoring must be simple, smart and must give advice.
- Monitoring in wards cannot be “fully continuous”, i.e. measurements may be continuous but there may be gaps in the data due to motion or due to the use model.
- Monitoring in wards cannot assume that the user fully understands the implications of the findings.
- There must be a kind of central monitoring system and administrative system that is optimized for the ward use.
- This system must be integrated in the care processes and hospital information system
- Must be unobtrusive, user friendly, avoid skin irritation, be compatible with nurse workflow, preferably reduces workload and be of low cost.
- For wearable devices size, weight, form factor, body location, battery lifetime (more than 24h, preferably > 3 days) and connectivity are crucial for successful application.

10.2.4.1 Spot check monitor system

The first approach was to introduce lower cost small spot-check monitors that are designed or optimized for use in the general ward, are mounted on a small cart and are optimized for an early warning system. An example is the MP5SC from Philips (see right picture of [Figure 10-21](#)). The advantage is that measurement data is more accurate and can be transferred automatically to a hospital or ward information system. Furthermore the EWS score is calculated automatically. Escalation can be automatic and data is stored and archived in the information systems of the hospitals.



Figure 10-21 Nurse based spot check system which uses a dedicated spot check monitor that also calculates the EWS score.

The data generated by the ward measurements must be translated into an actionable form and become available to clinicians. Data must also be stored in the electronic patient record (EPR) and hospital information system (HIS). An example of central monitoring environment for the ward is shown in *Figure 10-22*.

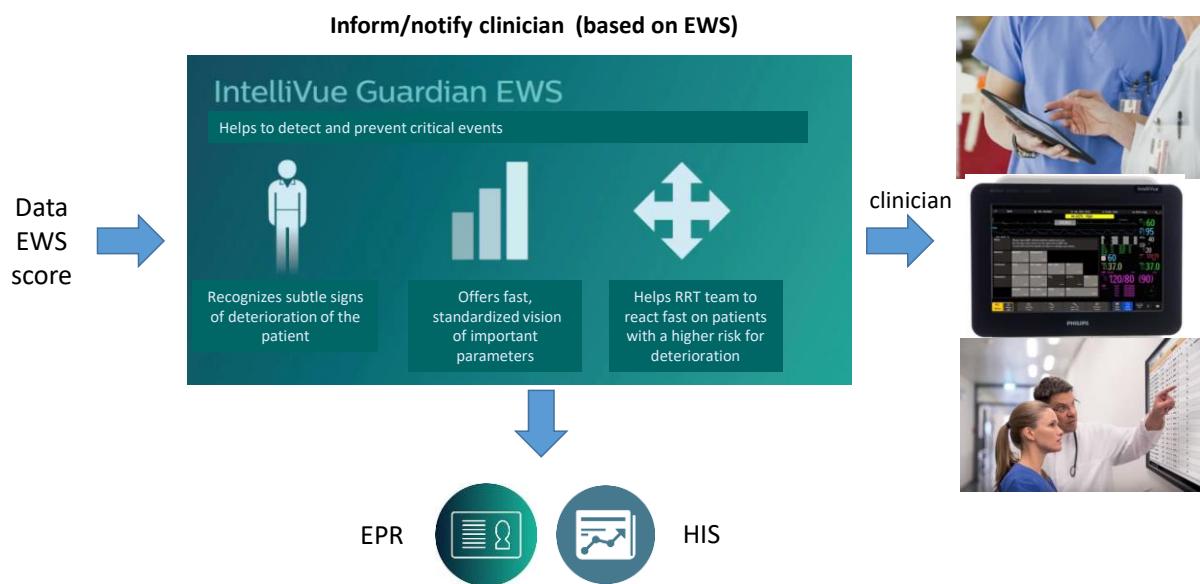


Figure 10-22 Diagram of the IntelliVue Guardian EWS system for the ward.

The data from the MP5SC spot check is entered via a wired or wireless link to the server with the Guardian software. This server is coupled to the hospital network. The user can configure the early warning scoring system and after the data is entered the EWS score is calculated, warnings can be generated and supporting information (also from other sources) can be shown to the clinician(s). Furthermore notifications can be sent to team members and the EWS process is guided. A standardized visual presentation of data is available. Data can be sent and displayed on monitors,

smart devices or on large displays in the nurse station in the form of dashboards or lists of the patients in the ward with indication of health status and trends. It is clear to the nurses which patients need most attention.

The above system has drawbacks, it requires major investments in infrastructure and the interval between the monitoring periods is still too long and events can be missed. A more continuous monitoring system is needed for a better coverage of trends in the patient disease state. Some options are described in the following sections.

10.2.4.2 Wearable devices

Most of the vital sign parameters can be measured using small devices attached to the patient and that can travel with the patient. Small transport monitors or telemetry devices have been proposed but size, weight and cost are too large. Philips has designed measurements to reduce cable clutter in the high acuity area but the devices can also be used in the ward. These are the Philips Cableless measurements. They are shown in *Figure 10-23*. Pulse oximetry, non-invasive oscillometric blood pressure, respiratory rate (RR) and heart rate (HR) (RR and HR are measured with an accelerometer). The accelerometer can also be used for posture detection, fall detection and activity monitoring. The devices can be connected to a monitor via a wireless proprietary module that uses the short range ZigBee standard (2.45 GHz band). The data can be transferred via the monitor or a proprietary access point to the hospital network. The advantage of ZigBee is that it is mesh network which enables data transfer from device to device. The advantage of this so-called cableless system is that it is a patient-worn system that can travel with the patient. The SpO₂ and RR and HR measurement are continuous but due to motion artifacts data can be corrupted. Especially the RR and HR pod is susceptible for motion artifacts as it is a motion sensor. This is in principle not a problem as continuous measurements are not required, the measurement frequency is much higher than the spot check frequency.



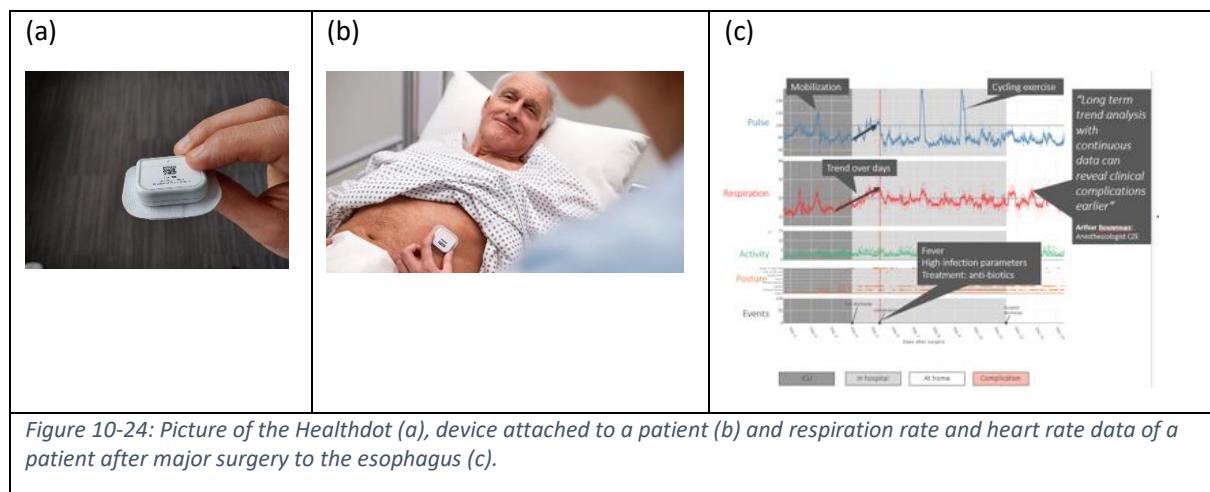
Figure 10-23 Philips Cableless measurements attached to a patient. They include heart rate, respiration rate, pulse oximetry and oscillometric blood pressure.

This was the first system of wearable devices that was commercially introduced in the market. Some valuable lessons were learned. High overhead costs of the wireless infrastructure, too high device costs, procedures, short battery run time (< 1 day), battery management (recharging, coupling to a patient ID), device loss (RR pod), non-optimal ease of use, and the non-optimal fit in the ward workflow limited the use of this system. The most important parameters of the EWS system are RR and HR. These can be measured with smaller devices which are optimized for ward use, some options are described next.

10.2.4.3 The HealthDot

This is an improved version of the device used for the cableless patient. The device has been redesigned by Philips Design and technology and Holst/IMEC for a use without a hub. It is optimized for a long battery run times of at least 2 weeks. Modern lower power MCU devices, a new radio standard designed for IoT (LoRa) and firmware were optimized for the use case. This device was introduced in the Netherlands in 2020 and was evaluated in the Catharina Hospital in Eindhoven on peri-operative care patients. The device is shown in *Figure 10-24*. It is a small sized box which must be attached below the left rib cage. The device is equipped with a LoRa radio, a radio standard designed for low power internet-of-things applications. The device and radio link is started by pushing a small knob, no hub is needed. The data is sent via a network provider (KPN) to a cloud based storage system where further processing can be done. A Zinc-Air battery is used for electrical power. A battery run time of 2 weeks has been demonstrated, a 30 day use period is possible after further optimization. Advantages of this device are:

1. The use model is set and forget, a single action starts operation.
2. No hub is needed, this is a large advantage.
3. Respiratory rate, heart rate and activity can be measured
4. Data is sent automatically to the cloud and is available via a standard internet link for clinical evaluation and monitoring of health state
5. The long battery run time allows many use cases including a complete peri-operative tract from home to hospital to home
6. The device is well suited for ward and home monitoring
7. The device is not obtrusive and does not limit user activities such as showering.
8. Pre-surgery base line data are available, this data can be valuable during treatment
9. Length of stay in the hospital can be reduced with several days
10. The device cost is low and it can be used as a disposable



In *Figure 10-24* (c) a time series of respiration, heart rate and activity is shown for a patient after a severe surgery to the esophagus. Initially after surgery a trend of increasing respiratory rate was observed which was related to an infection, after an antibiotic cure the patient recovered. The data also shows periods during exercise (high peaks in heart rate). Further larger studies on a patient group of bariatric surgery patients showed that length of stay in the hospital after surgery could be halved, resulting in a large reduction in costs, reduction in the infection rate and improved patient comfort. Note that the LoRa radio system is not available in all countries, there is an initiative that

LoRa is available in all countries of the EU. A LoRa gateway can be placed in- and outside the hospital when there is no coverage of the LoRa WAN by a telecom provider. At present the Healthdot device is used in the Netherlands where it can also be used for home monitoring.

10.2.4.4 Smart Plasters

The respiratory rate and heart rate are measured in many bed-side monitors using an ECG measurement. The heart rate can be extracted from a short lead ECG measurement (electrode distance of a few centimeter) from the R-R peak interval. The respiratory rate can be determined from the small bio-impedance changes of the thorax impedance. An example of the Philips Patch/biosensor is shown in *Figure 10-25*. The sensor can be glued to the skin via the ECG contacts at a location on the chest near the arm pit. The size is approximately 10 cm by 2 cm. The weight is less than 30 grams. The sensor also contains an accelerometer (posture, activity, step count, fall detection), a thermistor for the skin temperature, a microcontroller and a short-range (10 m) Bluetooth low energy radio. The radio is connected via a hub type relay device to a server with the Guardian software. When the patient is separated from the relay by more than 10 meter the patient needs to carry the relay device. The rest of the system is similar to that of the Guardian system for the spot check and cableless measurements. The battery in the patch is not of a rechargeable type, the patch is a single-use disposable device. The materials of the patch and components are selected for waste management and disposal. The advantages are its form factor, ease of use, longer battery life (several days), avoidance of battery charging and use of an established standard for the wireless connectivity. The cost should be very low (~10 \$) as it is a disposable, otherwise the use model will not be financially attractive as very large numbers of devices are needed. The ECG electrodes contain aggressive chemicals and this may cause skin irritation. Furthermore the relay or hub is not an attractive feature, it adds to the cost and degrades the reliability, it would have been better to use the existing Wi-Fi network of the hospital or a new wireless telecommunication standard for internet-of-things. Furthermore in case of interference with Wi-Fi signals from the hospital network, the Bluetooth connectivity will be severely degraded, the Wi-Fi transmission power is much larger. There are more suppliers of patch like devices. This type of device is presently tested in hospitals and is used on a small scale.

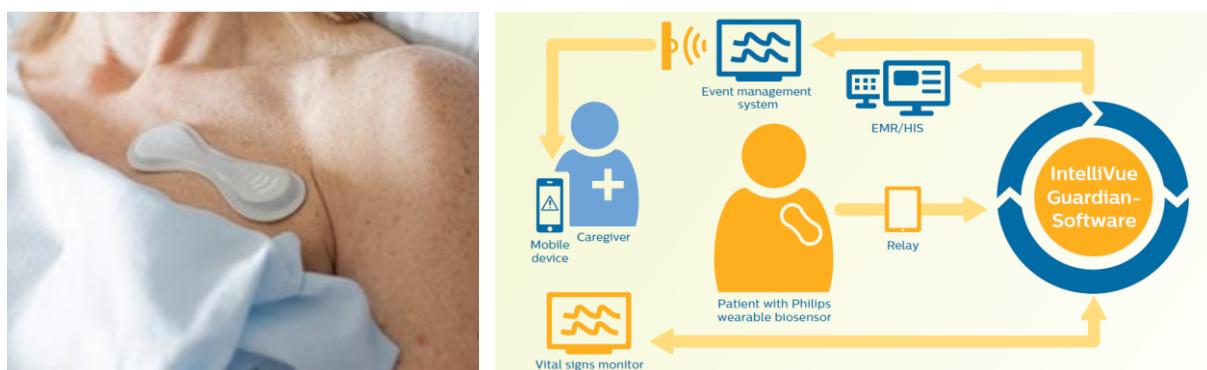


Figure 10-25 Philips biosensor patch and integration in hospital system.

10.2.4.5 Health watch

A health watch or smart watch measures heart rate and activity and can be used for patient monitoring in the general ward (see *Figure 10-26*). A reflective green ppg sensor is used for the pulse rate measurement at the wrist. The modulation envelope of the ppg signals contains information on the respiration rate and in ideal cases the respiration rate can be extracted. Other mechanisms (thermoregulation, central nervous system regulation of local blood flow) may also modulate the ppg signal and the accuracy of the extracted RR may be degraded. Furthermore the contact of the sensor

with the skin is a problem for some patients. An accelerometer provides similar functionality as for the patch device (fall and activity). For the connectivity Bluetooth and Wi-Fi may be used, for the access points similar solutions as used for the patch are proposed. The main drawback of this device is the lack of a robust and reliable respiration rate measurement. New consumer devices with increased functionality (ECG and SpO₂) have become available, the application area for such devices is still not clear, initial studies aim at screening in the home for atrial fibrillation and sleep apnea.



Figure 10-26 Philips health watch and optical sensing element.

10.2.4.6 Camera based vital sign monitoring

It is possible to extract the respiration rate and heart rate with a video camera. Measurements are non-contact and unobtrusive for the patient. The heart rate can be extracted by small variations in skin color due to changes in blood volume near the skin. The respiration rate can be extracted by measurement of chest and abdomen motion. A diagram of a camera based ward monitoring system is shown in *Figure 10-27*.



Figure 10-27 Schematic diagram of a camera based ward monitoring system.

The camera module is wall mounted or integrated in a bed-side display. Via advanced video processing techniques the heart rate (skin color changes) and respiration rate (chest motion) can be determined. Further options are detection of bed leave, falls and activity. The camera is linked to a specific room and bed, the patient cannot be followed when the patient is transported or is walking through the hospital. There is concern about privacy and extraction of the vital sign parameters may be difficult in some conditions. Power consumption is relatively large and powering will be done

mostly from the mains supply. There are wired and wireless connectivity options. The cost of the module are acceptable but larger investments in infrastructure are needed. The technology is not mature yet and is less suited for mobile patients. A significant research effort is still required before this method can be used in the hospital.

10.2.4.7 Smart Bed

A piezo pressure sensor can be added to a bed and in principle heart rate, respiration rate and patient motion can be measured (see *Figure 10-28*).

The sensing element can be placed between the matress and the base of the bed. Patients are only monitored when in bed. This measurement is unobtrusive. When the patient is discharged beds are removed from the ward room for cleaning and removed to another location. This is a logistic problem. The sensor must also be cleaned.



Figure 10-28 Smart bed system for measurement of respiration and heart rate.

There are now many options for monitoring in the ward, it is still not clear which method is the best solution. The system, its costs, IT infrastructure and user-friendliness will ultimately be decisive. Furthermore the reduction of the high false warning rate is still a major issue. Integration in the hospital workflow is the second major issue. Monitoring in the ward is the new frontier of patient monitoring and is potentially a very large market. However the majority of the hospitals still rely on (manual) spot-check monitoring techniques. These hospitals will invest in new techniques only when there is the highest level of evidence that the system improves outcome. The status of the level of evidence of RRT systems is discussed in the next section.

10.2.5 Status of Rapid Response Systems

The cost-effectiveness and effect on patient outcomes of RRT systems have been questioned and remain to be determined. There is no level-one evidence (double blinded randomized study) to support for the introduction of RRT systems in a hospital. There has been one attempt (MERIT study in Australia) and the outcome did not support the study hypothesis that RRT systems improve outcome. It is extremely difficult to conduct such a study and there were several weaknesses in the study. It appeared that the statistical power was not sufficient to support improvements in outcome. The organization of such a test requires a much longer preparation period and system optimization and tuning period in hospitals where RRT was introduced. There were also several confounding effects. It appeared that hospitals in the reference group improved the quality of care during the

study. Most evidence supporting the effectiveness of rapid-response systems comes from un-blinded, nonrandomized, short-term studies at single centers, in which outcomes before and after the implementation of such systems were compared. Further evidence comes from meta-analysis studies of existing single center studies. An example of review and a meta-analysis of data is shown in Figure 10-29 (de Jong et al. [2]).

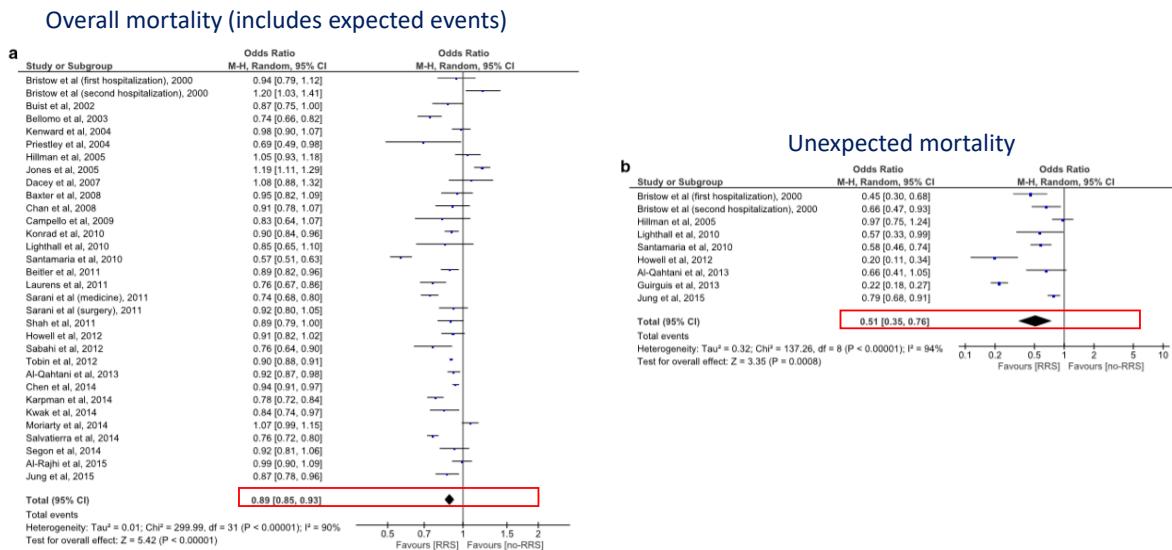


Figure 10-29 Results from a review and meta-analysis of single center studies. (From de Jong et al. [2]).

It appears that there are statistically relevant improvements in overall mortality and unexpected mortality. Many studies support the RRT system but a large multi-center randomized study with clear level I evidence to support the RRT system is needed to convince the majority of hospitals before there will be general acceptance of the system.

Despite this lack of evidence there are more than 3500 hospitals in the USA where an RRT system is implemented. The system is implemented in most hospitals in Australia, New Zealand, The Netherlands and the Scandinavian Nordic countries. In most cases the afferent part of the RRT system still relies on spot-check monitoring, continuous ward monitoring is relatively rare. Note that it requires a long term and large effort and large long term investments before a good RRT system is present in a hospital. It is likely that the RRT system performs non-optimal in most of the hospitals. It takes much more time and effort than initially anticipated to introduce high quality RRT systems in hospitals.

10.3 Home Monitoring

There is a need to improve care of patients in the home with chronic illness, prevent exacerbations and reduce visits of the patients to (specialists in) the hospital. In the USA and Europe there is drive to reduce unplanned re-admission to the hospital after discharge from hospital. When in the USA a patient is admitted to the hospital within 30 days from hospital discharge, the hospital has to pay for the care after re-admission.

In this section the focus is on long duration (at least 7 and preferably 30 days or longer on a single battery charge) semi-continuous monitoring of clinical-quality vital signs monitoring for home use cases (abbreviation “continuous”). Use cases such spot monitoring for glucose monitoring, heart failure, Asthma, COPD and hypertension are not included¹⁶. Telemonitoring for cardiac arrhythmias

¹⁶ In these cases mostly intermittent monitoring is used.

and chest pain need continuous ECG monitoring, but do not require the longest battery life time¹⁷. They are not discussed here. Consumer type of devices (wrist worn) are excluded because of the quality of data and the short battery lifetime. Note that reimbursement is critical for wide scale. Note that remote monitoring of patients outside the traditional care settings like hospitals is even more complex than monitoring in the hospital ward. A diagram of a remote patient monitoring system is shown in Figure 10-30.

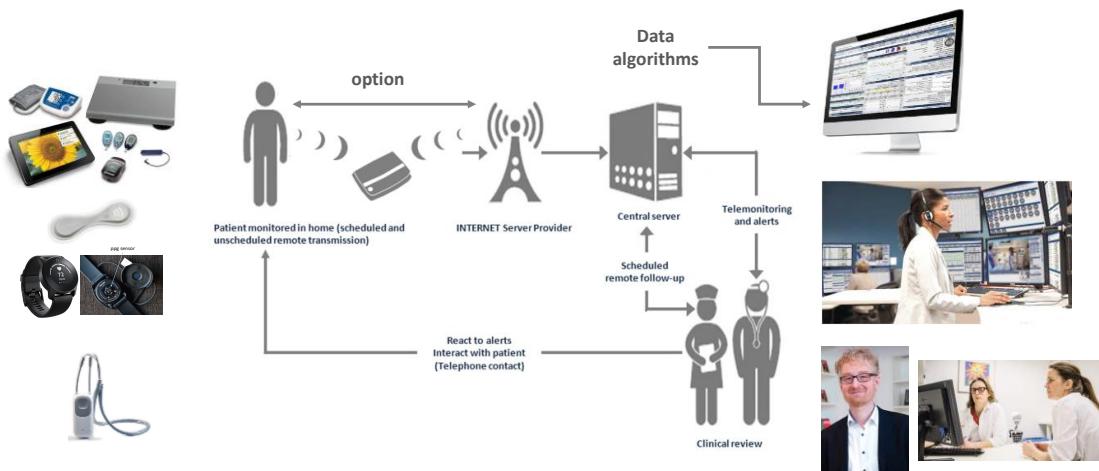


Figure 10-30 Diagram of remote patient monitoring system.

Devices for monitoring vital signs are wearables and are similar to the devices used in the ward. The consumer industry targets this market by introducing special functions in smart watches. For some parameters like blood pressure, and weight dedicated devices are needed. The battery run time of the wearable devices has to be much longer than for the ward devices as professional support may be needed to restart monitoring. A run time of 30 days would be desirable.

The on-body monitoring device is battery operated, has sufficient internal storage capacity and must have a secure wireless connectivity to a hub or bridge type device or directly via a mobile phone network (3G,4G,5G, IoT). Continuous transmission of data will drain the battery in a short time, therefore (selected) data will be sent in bursts to the hub. There are multiple potential solutions for a hub, the optimal solution is not clear. The system functions in the home environment and it is not sure if all components (internet, Wi-Fi, smart devices) are present and/or function well. A smartphone as a hub could be a candidate.

The use of consumer devices and dedicated clinical grade devices for home monitoring is discussed below.

10.3.1 Telemonitoring - Wearable Consumer Devices and Apps

Important requirements for such systems are: generation of actionable data, ease of use (especially important for the elderly patient group and clinicians), very long battery run time (> weeks), clear information and actionable steps for all users (both clinical and patients), and if needed easy communication with healthcare professionals. Major tech companies aim at the lucrative and large healthcare market. They offer a variety of health watches (see Figure 10-31).

¹⁷ See for instance the discussion on the Apple health watch



Figure 10-31 Health watches from Apple(a), Samsung (b), Omron (c) and Huawei (d).

The Apple watch has ECG, pleth, activity, wrist temperature, fall detection and SpO₂ (not medical grade) and FDA clearance for Afib detection using the pleth wave and ECG signal. It does not offer blood pressure and respiration rate. Its battery life is maximal 36 hours in low power mode. The SpO₂ data may not be used for clinical application. The Samsung watch includes a blood pressure estimation using pleth pulse wave analysis, it is not medical grade. The Omron watch uses a wrist cuff oscillometric blood pressure measurement with quality comparable to arm cuff measurements. Furthermore continuous heart rate (ppg based) and activity can be monitored. Battery life of 7 days is claimed for 8 blood pressure measurements per day. The Huawei health watch offer the same functionality as the Apple and Samsung watches and adds the same type of blood pressure measurement as the Omron device. None of the smart watches offer a clinical grade measurement of the respiration rate and SpO₂ level. The Omron device does provide a reliable spot check of blood pressure. Furthermore all devices use Android or IOS operating systems and standard wireless device communication. Battery life is too short for our use case (30 days), miss the most important measurement (respiration rate) and ease of use may not be optimal for the majority of patients. Another issue is privacy, it cannot be guaranteed that data cannot be accessed by these large companies and there may be back doors for data transmission (requirements by governments). In summary these devices may be well suited for their intended consumer use groups but are not suitable for many use cases of professional clinical grade monitoring in the home. These devices can provide useful information on the health state when vital signs are gathered over very long periods (years).

10.3.2 Professional wearable devices for hospital and home use

The COVID epidemic has accelerated the use of home monitoring and telehealth (see Figure 10-32). In the United States and Europe two applications for “continuous” home monitoring are readmission within 30 days after hospital discharge and monitoring during the perioperative care cycle. The driver is cost reduction for the hospital and improvement of patient care and outcomes. The most important vital signs for home monitoring and in general low acuity monitoring are heart rate, respiration rate and in some cases SpO₂. Furthermore activity and posture are important parameters. The heart rate and respiration rate measurements must be at least semi-continuous, i.e. there may be short periods (minutes) where data are classified as artifacts. Solutions must preferably be compatible with a thirty day measurement time on a single battery charge. This includes power consumption due to connectivity. The main connectivity technologies were discussed in a previous section of this chapter. Furthermore the sensor body location, its size and volume and adhesives used for the connection to the body are important. The most used wearable technologies use ECG and bioimpedance for heart rate and respiration rate. One ECG example and two others sensors using a different sensor type are discussed. Only devices with FDA or CE market approval are shown.

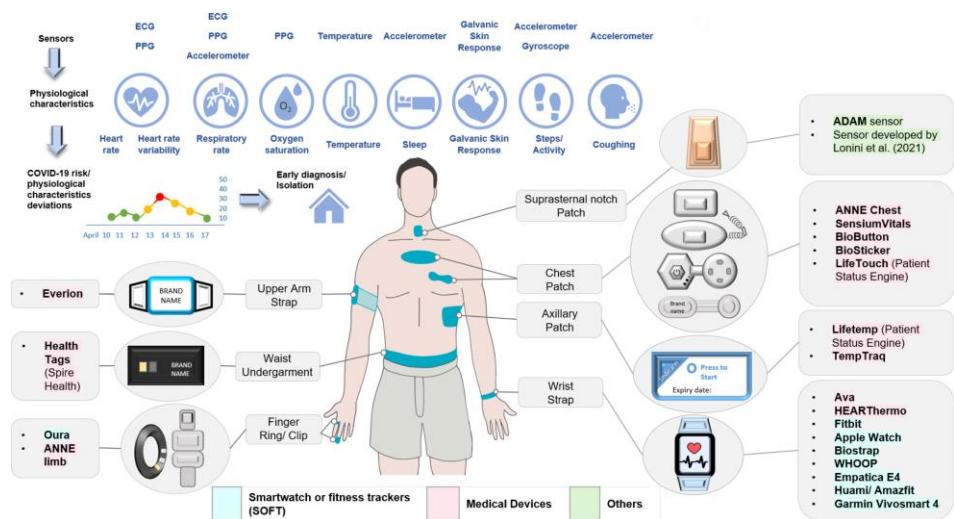


Figure 10-32 Overview of wearable devices and sensor locations used during the COVID epidemic.

10.3.2.1 Vital Connect – ECG based patch

The Vital Connect device is a disposable one-lead ECG patch device (Figure 10-33). A proprietary ASIC is used for the ECG and bio-impedance measurement. The ECG signal is used for heart rate and heart rate variability, bio-impedance is used to measure respiration rate. It also includes a 3-axis accelerometer for posture, activity and fall detection and a skin temperature sensor. It can be used for periods up to 7 days. The ECG signal is of good quality and can give additional information to the clinician. However it may be used for clinical decision making. The heart rate can be measured both in stationary and mobile conditions. Accuracy is roughly ± 5 beats per minute. The respiration rate accuracy is ± 3 breaths per minute. Up to 10 hours of data can be stored on the device. The wireless connectivity is via Bluetooth. The patient must be in the proximity (< 10 m) of the host device when continuous data is needed. A zinc-air battery is used as a power source. A remote service is available for monitoring and interpretation of the vital sign data.

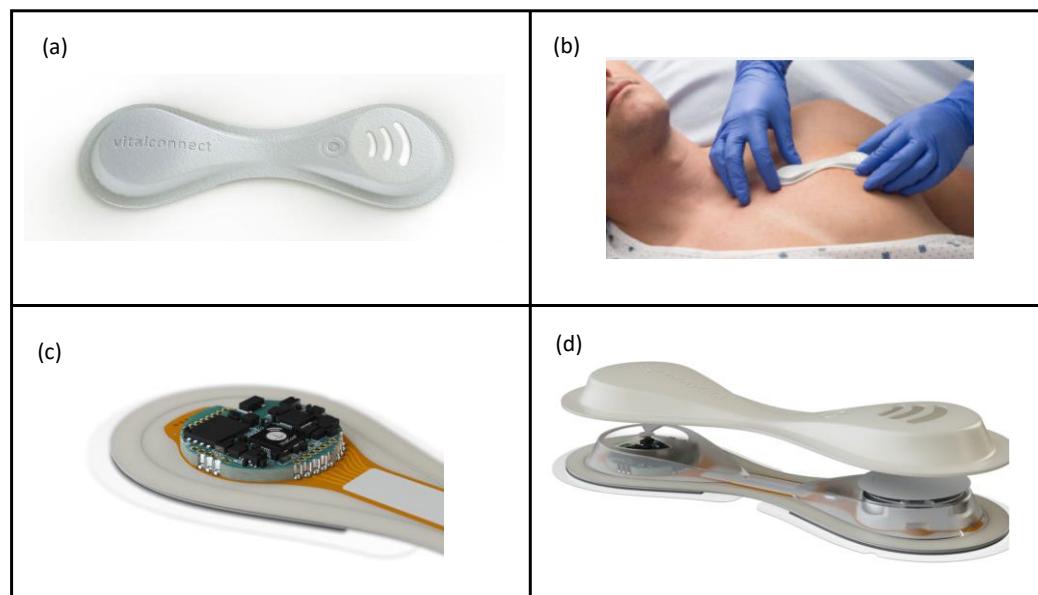


Figure 10-33 Vital Connect ECG patch wearable. (a) the patch, (b) body location, (c) View of ECG electronics and ASIC, (d) overview of interior.

10.3.2.2 BioIntelliSense Devices

This company uses heart and lung sounds to measure resting heart rate and breathing rate. There are two forms, the BioSticker disposable device and the BioButton rechargeable single-use device (Figure 10-34). The sensor measures the same sounds as used during a clinical examination with a stethoscope. The BioSticker consists of two compartments, the sensor part and the battery part. A 620mAh, Lithium coin cell (3V) is used, a battery run time of up to 90 days is possible, 30 days seems more realistic. The body position is two inches below the collar bone. Bluetooth is used for short range connectivity to different type of proprietary hubs including WiFi and 5G cellular. The hubs will automatically pair with the sensor when detected after turn on, no user action is required. The patient needs to be within 10 meters of the hub for data transfer. The solution includes cloud based system for data analytics and algorithmic-based alerting to enable actionable clinical triage and workflow efficiencies. A smaller and rechargeable version of the device is the BioButton. It form and size resembles that of the health dot. The lithium coin cell is replaced by a Li-Ion battery with a capacity of 200 mAh. A run time of 14 days is claimed. The battery can be recharged and therefore other use models are possible and re-use is possible. Note that heart and lung sounds are used. These measurements become invalid during motion but can also be compromised by speech and other disturbing sounds as for instance TV audio.



Figure 10-34 BioSticker and BioButton wearable devices including a hub for 5G data connection.

10.3.2.3 Philips HealthDot

About 15 years ago Philips research studied the use accelerometers to measure respiration and heart rates. This was very successful and a resp-pod module was developed for the cableless sensor family and introduced in the market around 2010. Since all wearable devices described in this chapter include accelerometers one might wonder why more complex measurements are used in the wearables. Note that wearable devices which use accelerometers for monitoring respiratory and heart rates have been clinically validated many times, it seems logical to use accelerometers as the only sensor for the measurement. This will drive down both power consumption and reduces cost. The Philips Healthdot was described in section 10.3.2.3 for use in low-acuity ward monitoring. Presently (2022) the Healthdot has been used both in the home and hospital situation and users found that it is the simplest clinical grade wearable device they have used. It is set-and-forget. After power on the device automatically connected to the LoRaWAN network of KPN. Note that the cost of the hub is removed from the system. The device has been in use in the Catharina Hospital n

Eindhoven, the users are very satisfied both on the reliability of the data and the ease of use. A cellular 4G-5G option is being studied, this would resolve the dependence on the LoRa standard which is not supported on a scale as the cellular solution.

10.4 Questions

The questions refer to both high acuity and low acuity monitoring

1. What is a shock, which types of shock can be distinguished?
2. What are the main changes in vital sign parameters for a septic shock?
3. What is the fluid status of a patient?
4. Describe the physiology of the main methods to detect a low fluid status.
5. Can the pulse pressure variation method be used for non-ventilated patients?
6. What is a scoring system? Why are they used?
7. What are the drawbacks of a scoring system?
8. What are the differences between low and high acuity monitoring systems.
9. What is a rapid response team, describe the main components of such a system.
10. What are the most important vital sign parameters for low acuity monitoring?
11. What is an early warning system?
12. What are the differences in requirements between low and high acuity monitoring systems?
13. What are the differences in requirements between low acuity hospital systems and home monitoring systems.

11 Patient Monitoring Technologies - Future Developments

Healthcare systems are facing tremendous challenges. The increasing number of patients, shortage of staff and rising costs of healthcare require major changes in the way that care is delivered. To maintain and improve the quality of care "new" technologies are proposed. In this section the focus is on the developed countries.

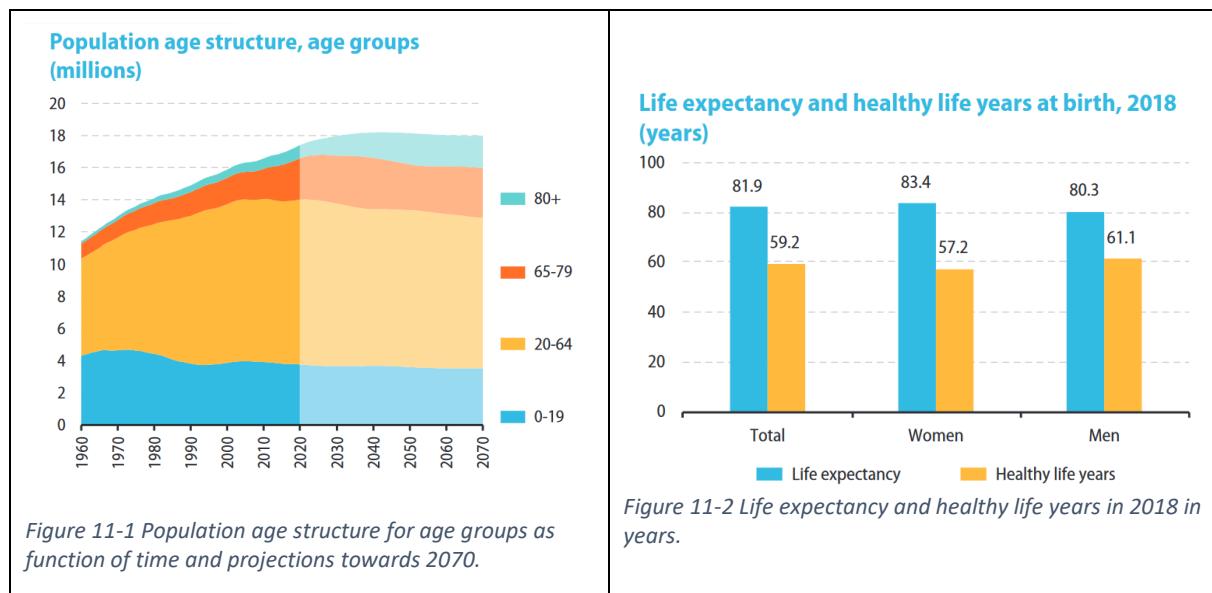
Present healthcare systems and trends are briefly described. The remainder of the section will focus on new patient monitoring systems. First the status of present patient monitoring technologies is reviewed. Limitations of the present monitoring systems are discussed. New clinical care models require major changes in architecture of monitoring systems, and a new generation of monitoring devices and systems is needed. Connectivity, miniaturization, telehealth and increased use of data and data analytics and artificial intelligence have been mentioned as key for future systems. Options for improved utilization of data, personalization of care and clinical support systems are discussed. There are a lot of strong opinions about the future of healthcare, especially from the large tech companies and their accomplices. It is our aim to give the reader a short overview of proposed options and add a personal view.

11.1 Trends in Healthcare Systems

At present there are major challenges to maintain or improve the high quality of care of healthcare systems. The main trends that affect healthcare are: the ageing population, increases in the number of people with (multiple) chronic diseases, obesity, increases in cost of care and shortage of (experienced) staff combined with overload of staff. Communicable diseases such as COVID further increase the strain on the healthcare systems. Note that the amount of medical knowledge increases exponentially with time. For a single specialization (like cardiology, oncology etc.) it is not possible to read the majority of new publications even if 100% of the available time is used. This increase in knowledge further aggravates the problems. Major changes in the system are needed to cope with these challenges. New "exponential" technologies will play an important role to cope with these challenges. There is a clear clinical and societal need for solutions. In the following section trends in healthcare are illustrated with data from the Netherlands and Europe, in some cases data from the United States are added.

11.1.1 Ageing Population

The proportion of elderly increases with time in the developed nations. Data for the Netherlands and projections to the future are shown in Figure 11-1. The proportion of older people (65+) in the population has risen sharply over the past 60 years, this trend is expected to continue for the next 30 years. In addition, life expectancy after birth continues to increase, currently the life expectancy for males is more than 80 years. The number of years in which the health state deteriorates is increasing and is currently more than 20 years (see figure 11-2). During this period the most use is made of health care systems. Note that there is a large difference (~5 years) in life expectancy and years of good health between the highly educated and the lower income groups. Hence there are more elderly people and more years in which the health condition declines. As a result, the already substantial costs of healthcare will rise faster than the national income. Note that there are also fewer people in the age group of 20-64 years, indicating that the staff shortage in care will increase with time.



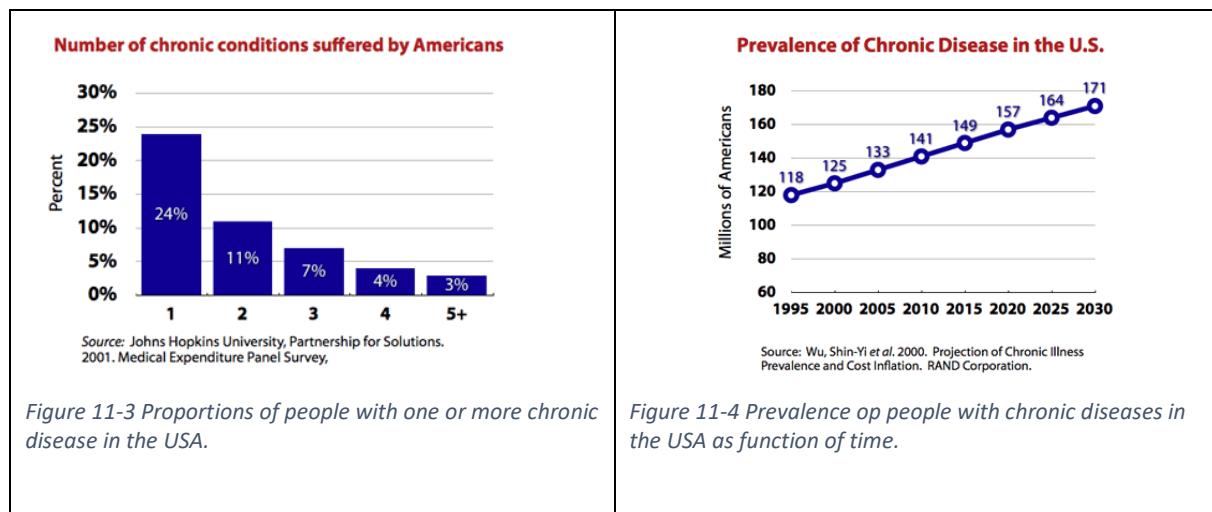
11.1.2 Proportion of Chronic Diseases

In the developed countries the proportion of the population with one or more chronic diseases has increased with time and is expected to increase further. Infectious diseases were a main cause of death in the beginning of the 20th century, at present mortality from chronic illnesses such as heart disease, cancer, pulmonary diseases, mental disorders, diabetes and stroke are the main causes of death. In the USA about 130 million people have one or more chronic disease (see Figure 11-3 and Figure 11-4) . According to a 2004 report: “*In 2004, almost half of all Americans, or 133 million people, live with a chronic condition. ... People with chronic conditions account for 83 percent of health care spending and those with five or more chronic conditions have an average of almost fifteen physician visits and fill over 50 prescriptions in a year.*” The number of people with chronic diseases in the USA and the developed countries still increases and this trend is expected to continue (see Figure 11-4).

Lifestyle (sedentary lifestyle, processed food, smoking, ...), pollution (air quality, industry waste) and obesity have been named as important factors for the high proportion of chronic illness. Recently it has been proposed that there are more factors of importance such as bacterial and virus infections.

A Dutch survey of 212,902 patient records found the following: 37% of the Dutch population older than 55 years had chronic diseases, in this group 67% had more than one chronic disease. Another study showed that the prevalence of chronic diseases doubled between 1985 and 2005. The proportion of patients with four or more chronic diseases increased by 300%.

In developed countries the proportion of people with one or more chronic diseases is very large and still increases with time. This causes an increasing strain on healthcare systems. From the previous sections it is clear that there is a strong increase in the duration and amount of care that has to be administrated and that the complexity of the disease state increases. Furthermore the number of people that require care increases. This raises questions if there is a sufficient amount of qualified clinical staff. This will be discussed next.



11.1.3 Shortage of Staff

In the Netherlands, approximately 1.4 million people worked in the care and welfare sector in 2020. This concerns employees and self-employed persons with a main job in healthcare institutions. This number includes the clinicians of all levels and all staff with an administrative or managerial position. About 1 in 6 workers is active in the healthcare domain. Care and welfare is therefore one of the largest employers in the Netherlands. At the same time, the healthcare sector is struggling with staff shortages and the demand for adequately trained healthcare personnel is only expected to increase in the coming years. The trends shown in the previous sections show that the load of the healthcare systems increases. On top of this there is the COVID-19 pandemic which has resulted in a strong increase in workload at all levels. The strain on the clinical staff has been enormous and this resulted in an increase in stress, burnouts and outflow of the highly skilled hospital staff. Furthermore more complex care and new technology is entering the field. This causes another increase in workload (rather than a reduction as is claimed by the technology developers). The shortage in highly skilled hospital staff is worrying. Moreover the average level of expertise is decreasing. It takes many years before the required expertise is obtained. Finally the complexity of the care process increases and the amount of medical knowledge increases with a high rate. This increase is discussed in the next section.

11.1.4 Growth in Medical Knowledge

In Figure 11-5 the increase in the number of published manuscripts in the healthcare domain is plotted versus time. The number of articles grows exponentially with time. It appears that medical knowledge grows in time with an exponential rate. In the 1980's it took 7 years to double the amount of knowledge, in 2020 this has decreased to slightly more than 2 months. This rate of change is difficult to follow for (individual) clinicians. Tools need to be developed to assist clinicians in bridging the knowledge gap. Thus on top of the challenges discussed in previous sections comes the staggering progress in knowledge and how to deal with this.

The increase in workload, amount of clinical knowledge and the shortage of experienced clinicians poses a formidable challenge for the healthcare system. Fortunately there are new "exponential" technologies that possibly may come to help. These technology options are discussed in one of the following sessions. The ageing population, the increase in chronic diseases, increases in labor costs, increases in cost of new medicines and large costs related to the introduction of new technologies and IT systems raise the question if the future cost of healthcare is still affordable, this is the topic of the next section.

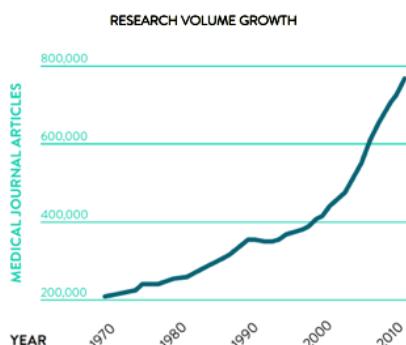


Figure 11-5 Number of journal articles versus year

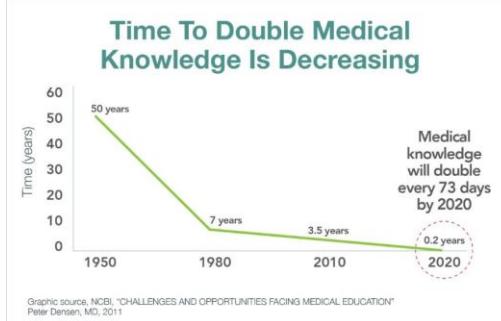


Figure 11-6 Time to double medical knowledge versus year.

11.1.5 Cost of Healthcare

Data from the EU and Dutch CBS (Central Bureau of Statistics) of the Netherlands are used in this section. The gross domestic product (GDP) and the cost of healthcare are plotted for the period 1998 to 2022 in Figure 11-7. The cost of healthcare is calculated in two manners, the blue curve is the Dutch CBS standard, the grey curve is the international standard which is used by the EU to compare spending in different countries. In the Dutch method more types of care are included (for instance elderly nursing home care). The numbers are corrected for inflation. The GDP grows with a factor of two in this period, the cost of healthcare with a factor of three. The fraction of the healthcare spendings of the GDP increases with time. There are abrupt changes in the increase of the GDP in 2008 and 2020 due to the financial crises and the Covid-19 pandemic. In Figure 11-8 the fraction of the healthcare spending of the GDP is shown for the period 1998 to 2022. There are periods with a steep increase in percentage of GDP spending followed by an almost constant proportion of the GDP spending. The rising costs were a problem, the flat parts are due to cost reduction measures by the government. The sharp increases were due to measures to improve the quality of care. For instance large investments were needed to shorten excessive waiting periods. From 2010 to 2019 about 10% of the GDP was spent on the healthcare system, this is a large proportion of the total GDP. Therefore cost reduction was key, the Dutch system was optimized for efficiency. This resulted in a limited capacity of care with little room for extra care. During the Covid period it appeared that the system was not able to cope with the increased demand, there was no margin to increase the number of patients at higher levels of care. As a result treatment of non-COVID related care was strongly reduced and postponed. This shortcoming of the system led to new and very large investments in personnel and materials. The strong increase in the last few years is excessive and experts warn that this trend might not be sustainable. Drastic measures are needed.

The question might arise if the degree of spending is related to the quality of care. This is illustrated in Figure 11-9 where the cost per capita is plotted versus life expectancy. There is no clear trend, for the same life expectancy the costs per capita differ a lot. For instance Japan has the highest life expectancy with a relatively low cost per capita. Note that Japan has the largest fraction of elderly and one would expect high costs per capita. On the other hand the United States has by far the highest spending but lags the group of highly developed countries in life expectancy. Note that in the USA 17% of the GDP is used for spending on healthcare. The degree of spending is weakly related to the quality of care. Data from the USA show that more spending is not always better. Note that the

healthcare expenditure in the Netherlands in 2011 was more than 3500 \$ per capita, the Dutch data point would be above that of Norway. In 2020 the Dutch spending per capita is 6000 Euro.

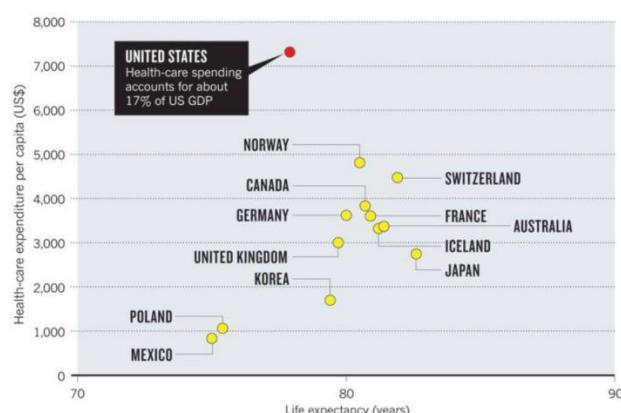
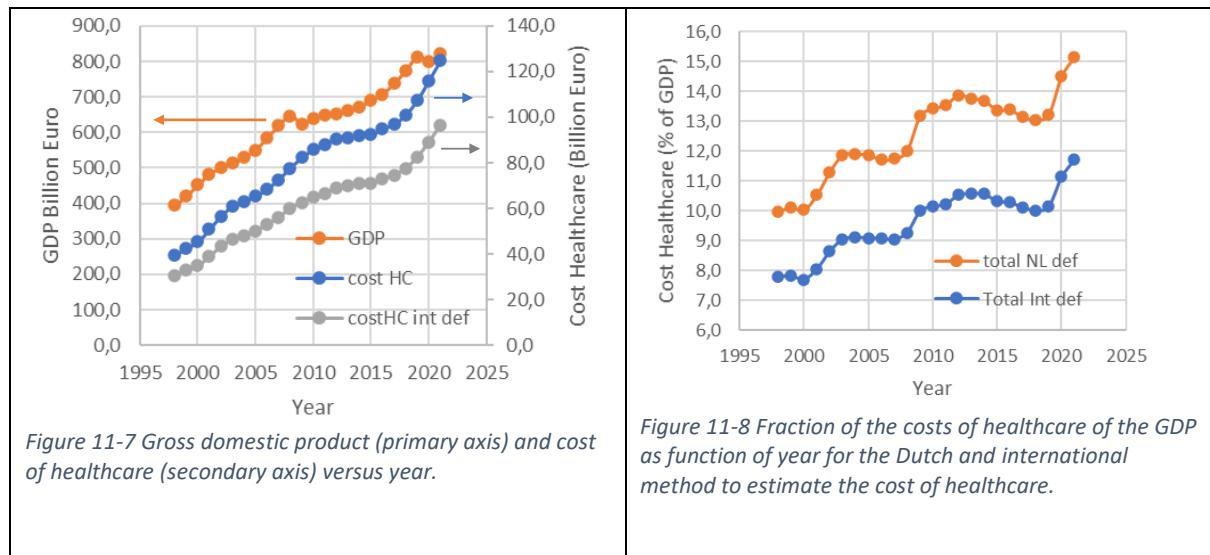


Figure 11-9 Healthcare expenditure per person per year versus life expectancy.

In most countries the healthcare system is under pressure due to the factors discussed above. The number of patients increases, there is a shortage in experienced clinicians, the workload increases, the strain on the staff is large. The cost of the healthcare system grows at a rate faster than the GDP, the end of this growth is not in sight. Something has to change. As said before many experts and companies believe that new exponential technologies come to help. This is discussed in one of the following sections. The future will learn if these claims are realistic. In the remainder the focus is on patient monitoring. Before going to a discussion of new technologies first the challenges and issues of the present monitoring technology are discussed.

11.2 Challenges and Issues with STATE-OF-THE-ART Monitoring Systems

Monitor architecture and technology has hardly changed in the last 20 years. Systems have become a little smaller, a little better, a little more efficient and a little more mobile and a little more expensive. Note that in the same period the computer and consumer technology has made an enormous progress both in increase in performance and in reduction of cost per function. Consumers are used to have access to information via the internet at any location and at any time. There is a large gap in technology between patient monitoring devices and these consumer technologies. Presently there are many different devices in use, each optimized for a specific department, specific

application, this causes inefficiencies during the patient trajectory in hospitals and loss of data. This leads to inefficiencies in the care process and degrades patient safety. This is illustrated with a an example below.

11.2.1 Transitions in monitoring during hospital care

The different phases in the care process and use of monitoring devices during a myocardial infarction (MI) event are shown in Figure 11-10.



Figure 11-10 Different stages in a myocardial infarction event and monitoring devices used in the different care settings.

A heart attack occurs in the home setting and the 112-emergency number is called. An ambulance is sent to the patient and EMS paramedics investigate the patient using a monitor-defibrillator. The preliminary diagnosis is myocardial infarction and the patient is transported in an ambulance to the emergency department of the hospital. During transport the patient is continuously monitored by the monitor-defibrillator, the EMS paramedic reacts to alarms. The monitoring data from the ambulance are not transmitted to the hospital, in most cases the hospital is warned via a dedicated radio system of a suspected MI event, the hospital has little further information. When the patient arrives in the emergency department the sensors and cables are removed by the paramedics and the monitor device is brought back to the ambulance. In most cases there is only verbal information exchange and as said before there is no transfer of monitoring data. The EMS staff starts almost from scratch. The EMS staff connects the patient to the EMS monitoring devices, this can take a few minutes and valuable time is lost. After investigations it is clear that a large coronary is blocked and it is decided that the patient must be sent immediately to the Cathlab to open the blocked coronary artery. Again, the sensors and cables are removed, valuable time is lost, and monitor data is not available for the team of the Cathlab. Again the monitoring devices in the Cathlab need to be connected and so on. After stent placement the patient is brought to the coronary care unit for observation and a telemetry device is connected to the patient. The wireless connection needs to be made. Note that a loss in time and data happens at each transition. Furthermore note that each department uses specific devices optimized for the use in the department. This is not efficient and requires a large overhead in the storage and use of the devices. Finally the patient is brought to the ward, there is only spot check monitoring, typically every 8 hours. In these period a lot can happen and studies have shown that this poses risks to the patient. When the patient is discharged from the hospital, the patient is not monitored during the first weeks. Again in the home situation severe degradation of the health state occurs regularly in the first 30 days and readmission in the hospital is common.

In summary there are many disruptions in the monitoring process, each department has own systems, systems do not travel from department to department and monitor data gets lost. There are many transitions that increase workload, there are gaps in monitoring, valuable time is lost, data is missing, and as a result the clinical care might not be optimal and patient safety is degraded. There

is a need for a new monitoring architecture and infra structure that solves these problems. There is a demand for continuous monitoring where data are sent to a central storage system where it can be viewed at any time from everywhere. Preferably a single device that can travel with the patient, that can be adapted to the level of care and that is compatible with mobile patients is desired by the users.

11.2.2 Gaps in Monitoring

Patients in the high-acuity departments of the hospital are well monitored, the patient-to-nurse ratio is small (1:1, 1:2) and follow up treatment is organized well. However as described in the previous section when there is a transition in the level of care there are gaps of several minutes where there is no monitoring. In the low acuity departments the situation is worse. For instance in the ward there is practically no vital sign monitoring. As described in a previous chapter many unanticipated events and potentially life-threatening conditions occur in the wards (see chapter 10.2). Note that many of the in-hospital cardiac arrests and cases of respiratory depression occur in the ward and they are often unnoticed by the clinical staff. In the ward the patient-to-nurse ratio is around 1 to 10, nurses do not have sufficient time for individual patients and the workload is extreme. Nurses in the ward have no training in the use of monitors. Monitoring in the ward is very complex, the health status is variable, patients are mobile and their location is often not known. New monitor devices for such use cases must be patient-worn and wireless connectivity and power consumption are major challenges. Furthermore, there is no optimal solution like a central station for the ward. Alarming will be even a greater challenge than for the ICU and OR departments. Presently there is no optimal solution for monitoring at low acuity environments like the ward and home. Preferably the new technology of the future should be scalable, i.e. it can be used both in the high acuity and low acuity environments. The device should be small enough such that it can also be patient worn. For such a future technology there will be no monitoring gaps anymore. Later we will discuss embodiments for such a technology.

11.2.3 Patient Worn Devices

To resolve the gaps in monitoring and improve patient safety a large fraction of patients outside the ICU and OR should be monitored on a regular basis. A continuous measurement is not needed as long as the measurement frequency is much higher than that of the spot checks (minute intervals are good enough). Often patients are mobile, therefore patient worn devices are a better solution than devices that can only be used at the bedside. Telemetry devices are not optimal for this use case, they are too large, too expensive and are obtrusive for the patient, furthermore short battery life is a challenge. The battery life should be larger than 3 days, the average length of stay on the ward. A new category of small body worn devices is available but size, cost, ease of use, power consumption, obtrusiveness and connectivity solutions are not optimal yet. Again, ideally the universal reusable device as was discussed in the previous section could solve these issues.

11.2.4 Warning and Alarming

Alarm fatigue is a major issue in the ICU and even more in the low acuity areas. This issue needs to be resolved without degrading the sensitivity of alarming. If there are no solutions this will limit the use of patient worn monitoring for mobile patients in the ward or in the home environment. There is no good solution to resolve the high warning rates but there are many options in both hardware and software solutions. Utilization of data, recognition of patterns and improved data analysis are options. Unfortunately legal issues need to be considered and may hinder introduction of new methods.

11.2.5 Connectivity

Presently wireless connectivity systems in the hospital are congested and quality of service of ISM networks is not warranted. Power consumption is a challenge for continuous transmission of data in battery operated systems. There are no easy solutions. The main technologies are short-range radios developed for the consumer domain and long-range wireless networks developed for Internet of Things (IoT) applications.

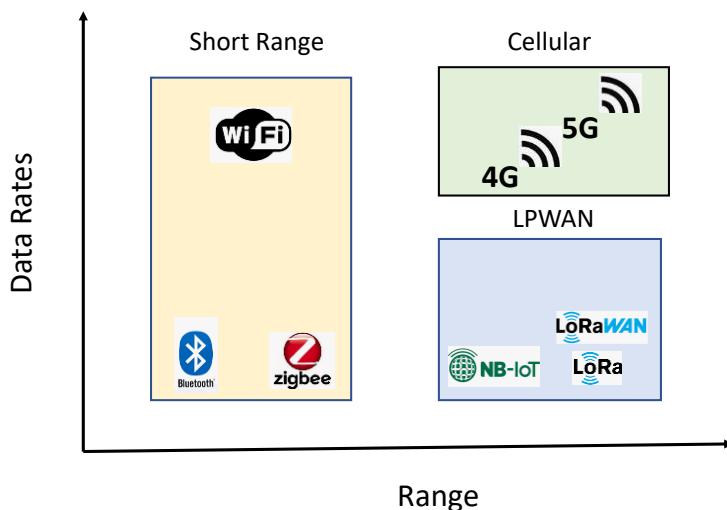


Figure 11-11 Communication standards for wireless sensor data transmission.

The main options are illustrated in Figure 11-11. The data rate for wearable vital sign monitoring applications is low (~100 Byte/s) much lower than these technologies can handle, therefore data transmission can be in the form of a burst mode. The power during transmission is relatively large for a wearable device (10-100 mW) but the average power can be sufficiently reduced (duty cycle 0.1%-10%). Note that clinical grade wireless transmission of data to a service provider is needed (i.e. quality of service and data confidentiality). Linking of the device to the host must be simple and fast (a few seconds) and must not be error prone. The user group (elderly, nurses) is not technically oriented, this is an important issue for the system design.

Several short-range (10m-100 m) wireless communication technologies are available, examples are Bluetooth, Bluetooth Low Energy (BLE), ZigBee (mainly used in home automation systems) and Wi-Fi (highest data rate but highest power). These standards either require a special hub or gateway connected to an internet provider or a host device (can be a laptop, smart phone, tablet etc.) with connection to an internet provider. This seems a good solution, these radios are available in almost all smart devices and the cost is very low. However smart devices have a short life span and the operating systems (Android, IOS) are updated regularly, specifications change. Some options needed by the device manufacturer may not be supported anymore. The use of such devices requires a major continuous effort for device developers mainly on software development and modifications in the hub or gateway. Data confidentiality and hacking are major issues that are not easy to solve. Furthermore ease of use is a concern because it cannot be assumed that the patient user group is familiar with this type of connectivity and connection to devices is prone to errors and can be time consuming.

Therefore long range (km's) wide area networks (WAN) are an interesting alternative. An external mobile service provider takes care of the connectivity to external servers and systems can become very simple. There are several low power wide area networks (LPWAN) already in use (see Figure

11-11). They are designed for low-power and low data rate applications including sensor networks (IoT). There are several options for long range, low power and reliable connectivity for Internet of Things (IoT). An example is LoRa. A service provider is needed for the WAN infrastructure (LoRaWAN) and the connection of the sensor to the server systems of the clinical users. After battery power is available the device will automatically connect to the LoRaWAN infrastructure. Push of one button is enough, it is that simple. This a very attractive option for wearable devices in the home environment. In the Netherlands KPN provides a reliable LoRa network infrastructure. A LoRa based sensor system can operate up to 15 years on two penlight batteries. In case of poor signal strength in a building a LoRa gateway can be used. However this standard is not supported on a sufficiently large scale by mobile service providers and this reduces the potential market size. A second LPWAN option is to use cellular technologies. The 4G standard is not optimal for low power IoT use cases. The cellular 5G standard was designed from the start to include options for low data rate and very low power IoT applications. The NB-IoT (Narrow Band IoT) and LTE-M (Long Term Evolution of Machines) standards are included in the 5G standard . The use of 5G cellular technology for IoT reduces the risk for long term connectivity support, almost all of the major mobile operators will support this system.

11.2.6 Cable Clutter

A major complaint from users is the clutter of cables when multiple measurements are needed (see Figure 11-12). By default, cables are cluttered, twisted and difficult to separate. This can lead to errors in measurements and even data loss. The ECG cables are the largest culprit. Cable clutter causes a lot irritation and extra work (patient movement for washing, position change, detachment of cables, cleaning of cables, etc.). Cable motion can cause multiple artifacts in the measurements. There is no good solution available despite many years of effort.



Figure 11-12 Clutter of cables for an ICU patient.

11.2.7 Utilization of Data

High-end monitors generate large amounts of waveform data and numerical parameter data. The data contains a lot of physiological information, both in the waveforms, variability, trends and in the time variation of vital sign parameters. This data is underutilized. This big data collection offers interesting opportunities for modeling, data analysis and artificial intelligence. An example could be early warning for hemodynamic instabilities, predictive alerts and meaningful alarms.

#signals level of processing	small	large
low	gives little information	garbage-in, garbage out
high	useless	

Figure 11-13 Matrix that relates the level of processing and the number of signals.

The options for further data analysis are illustrated in Figure 11-13. When the number of signals is small the analysis should be limited, there is little information available. When there are multiple signals available high-level processing and further analysis are possible. It should be noted that for all types of analysis the knowledge of the complete measurement system is important, for instance in the measurement certain features of the signal can be removed by analog and digital filtering and digital processing.

Improvements in monitoring technology, data utilization and data analysis are needed to provide actionable data and support that can be used during the complete care process. Furthermore new technologies have been proposed to improve quality of care and at the same time reduce costs of care. These topics are discussed in the next sections.

11.3 Technology Trends in Healthcare

There is no scarcity of forecasts of the future of healthcare technology, the major analysts and consulting companies are active in this area. A few examples of predictions are shown in Figure 11-14 to illustrate commonalities in the predictions of these organizations.

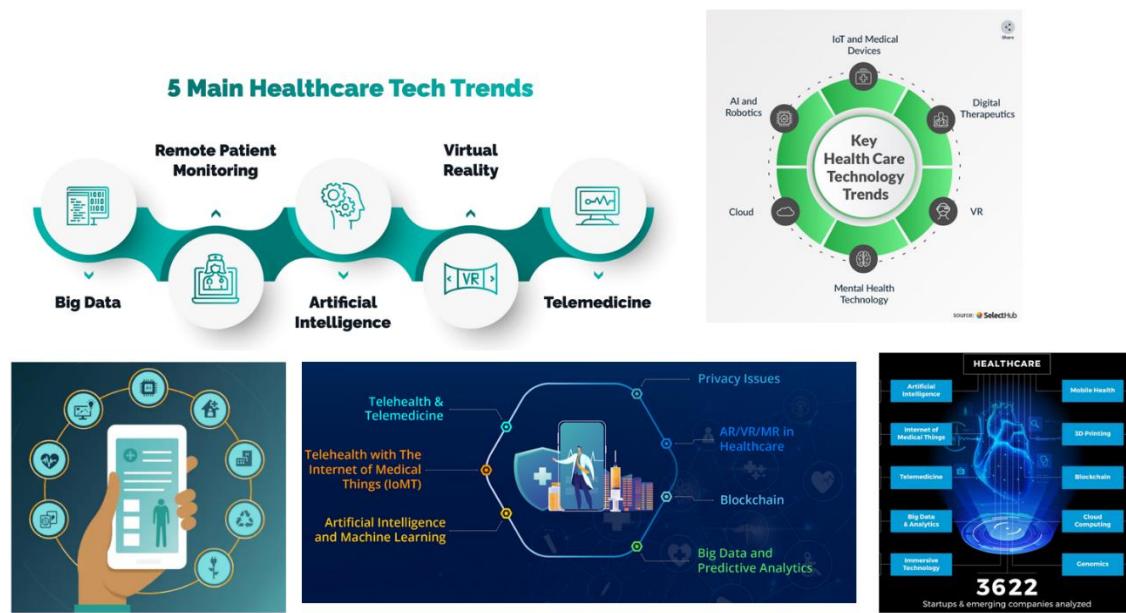


Figure 11-14 Examples of predictions of 2021 healthcare technology trends.

Wearables, apps, telehealth, monitoring everywhere, cloud based technology, big data and artificial intelligence are seen as the potential solutions for the main challenges of healthcare of the future. This is summarized in Figure 11-15. In the following sections we will give an overview of the main technologies and options and their status. Let's start with the measurements that are needed for

future monitoring systems including wearables and telehealth. Thereafter the other topics shown in this figure are discussed.

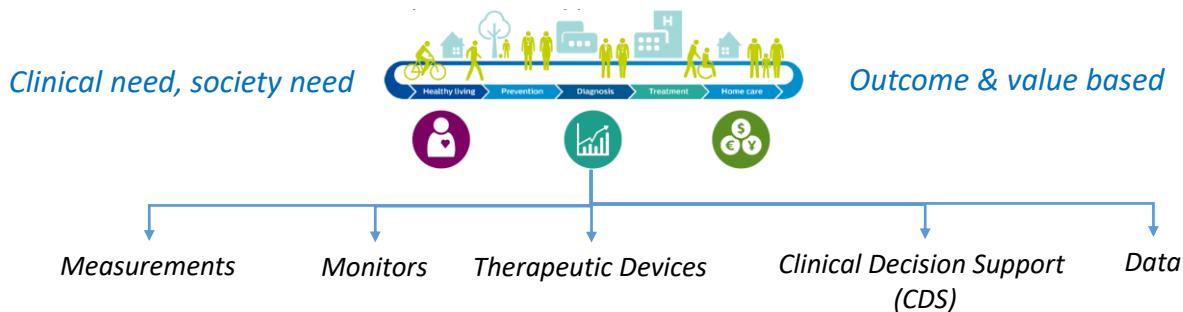


Figure 11-15 Continuum of care process and its requirements on monitoring technology.

11.4 Clinical Measurements

The technology status for the vital sign measurements are analyzed in this section. ECG and bio-impedance are discussed first.

11.4.1 ECG and Bioimpedance Measurements

A PCB with an ECG ASIC for an IntelliVue measurement for a 12-lead ECG and its wearable counterparts for low acuity care are shown in Figure 11-16. The PCB area of a 12 lead ECG measurement is relatively large (area \sim 60 cm 2). This is mainly due to the large number of ECG leads and related overhead and measures for the protection against high voltages. Regulatory requirements for patient safety are stringent. For instance leakage currents between the various measurement pins and wires must be in the μ A range for 1000V voltage differences. The protection against large transient voltage spikes and current pulses (defibrillator pulses and ESD) and the 50Hz power supply lines induced noise requires two type of protection devices and several stages of analog filtering, this consumes a large part of the PCB area. Another large chunk of the area is occupied by the devices that are needed for the generation of three or more power supply voltages needed for the internal components. Note that backward compatibility of the module requires compatibility with 3V, 5V, 12V and sometimes even higher supply voltages. Finally the isolation from mains voltage is done with a large isolation transformer (protect against 7kV pulses!) and data isolation requires relatively large (optical) devices. The ECG ASIC and MCU microcontroller use only a relatively small fraction of the PCB area. This module is designed for a professional in-hospital use with the highest clinical demands on the signals of a 12 lead ECG application. The quality of the waveform features is very important, a large bandwidth is required and small signal distortion is needed. Furthermore the ECG electrodes are located on pre-defined locations on the body with separations as large as 50 cm or more. This requires a complex ECG cable. The module also measures the vital signs of heart rate (R-peak separation) and respiration rate (Bio-impedance). Power consumption is in the order of 10's of milliwatts. An ECG measurement based on this module is not suited for wearable use cases. With present technology a clinical grade 5 lead ECG device for a multiparameter patient monitor can be made of the size of a credit card and thickness less than one centimeter when new cables and connectors are used and different methods for data and power transmission are used. Prototype devices have been made but both size and power consumption are still too large long duration wearable use (i.e. requirements of several weeks for home monitoring).

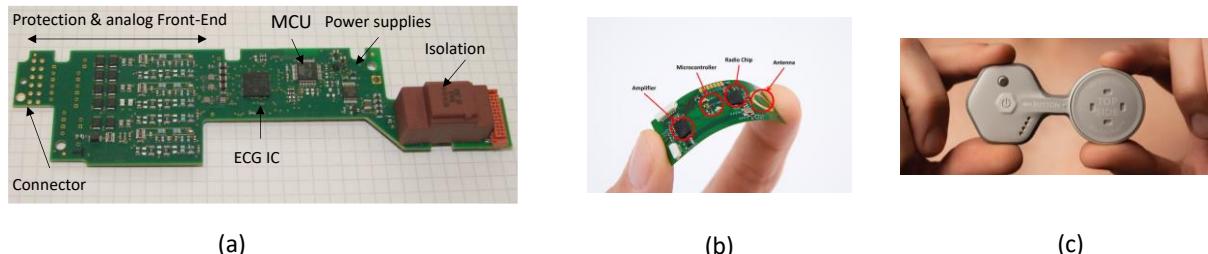


Figure 11-16 A 12 lead ECG module for an IntelliVue monitor (a), a flexible PCB from Holst-IMEC for a wearable short-lead ECG/respiratory rate module and a commercial product of a wearable for home monitoring using ECG and bio-impedance (c).

The requirements of a wearable device based on an ECG module differ from the in-hospital 12-lead ECG measurement. A one lead ECG measurement is used, protection against high voltage spikes is simpler. The heart rate is extracted from the R-R peak time difference, the ECG waveform is less important. Bio impedance is used for the respiration rate. Important requirements are long battery life of at least 3 days (preferably 30 days), small size and weight. Wireless connectivity is a must. Most ECG wearables use a simple ASIC for the 1 lead ECG signal and use a small ARM M3 or M4 MCU microcontroller with embedded short range radio for signal processing, feature extraction and connectivity. ECG Asics are available from large companies like TI, Analog Devices and Maxim. Furthermore there are proprietary ASIC devices from Philips and Holst/IMEC, i.e. the key components for wearables are available. The battery, the voltage generation components, the antenna and electrode size and their separation determine the size of the wearable. Data is transmitted in bursts to lower the average power consumption (the radio is the most power hungry device). For a 200mAh Li-polymer battery a 30 day run time has been reported. Wearable devices are often single use, high cost is an issue. Presently prices are in the order of 100 Euro or more, this limits one time use. Note that even for larger scale fabrication the bill of materials of the ECG-wearable are much higher than for other competitive technologies and this may limit its widespread use. For instance a device cost of 25 Euros for a single use device may not be acceptable.

11.4.2 Pulse Oximetry

There are many products for small size pulse oximetry systems and for devices for home monitoring. A distinction must be made for clinical grade measurements and consumer grade devices. A few examples are shown in Figure 11-17. The first pictures (from a to d) are clinical grade, the fifth (e) is a smart watch from Apple. The supplier warns that this device must not be used for medical applications. The location on the wrist is a bad location for a clinical grade SpO₂ measurement.

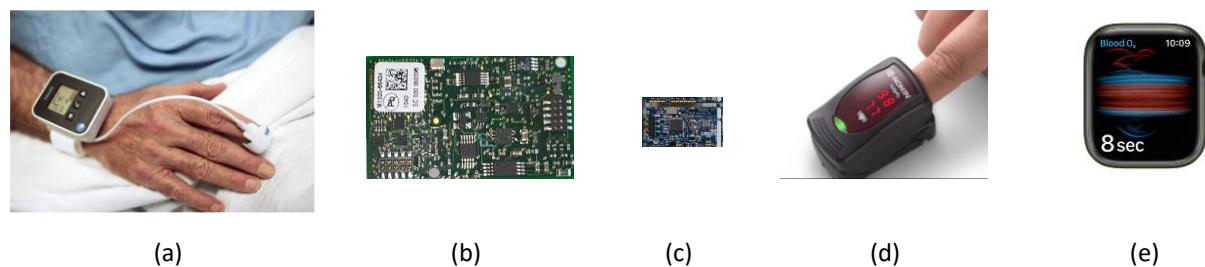


Figure 11-17 Wearable SpO₂ devices and PCB's are shown. (a) Philips cableless device, (b) PCB of the Philips device, (c) a PCB of a device with an SpO₂ ASIC, (d) a device from NONIN and (e) the Apple series 8 smart watch. The first four devices are clinical grade.

The requirements of the clinical devices in terms of signal to noise, dynamic range and signal processing are on the limit of what is technically possible. For a large signal to noise the light

intensity of the LEDs must be sufficiently high. This limits battery life for continuous use. It is already a challenge to use a battery operated device for more than a day. However continuous measurements of SpO₂ are not needed and intermittent measurements are good enough for the mobile patient category. In Figure 11-17 a wrist worn device from the Philips cableless patient monitoring solution is shown. The measurement is done on the fingertip. The measurement hardware of this device uses discrete components and is shown in Figure 11-17b. The size of clinical grade PCB's can be reduced by using ASICs. An example of such an embodiment with a ASIC IC from Texas Instruments is shown in Figure 11-17c. A product from NONIN for clinical grade home monitoring is shown in Figure 11-17d. Note that the quality of the measurement depends both on the hardware components, digital signal processing and to a large factor by the removal of artifacts by the post-processing algorithm. In summary there are clinical grade solutions for home and wearable SpO₂ monitoring. The use case is limited to intermittent monitoring due to the power consumption of the LED devices. Note that in wearables often a green ppg sensor is used for continuous measurements of the heart rate, a few days battery life is possible. Furthermore there are many studies on the extraction of the respiration rate from the ppg signals. It is suggested that breathing rate can be extracted from the modulation of the ppg signal. Note that respiration rate extracted from ppg is not used in clinical products because it is not clear if the data can be trusted in all use cases. For instance thermoregulation and regulation of blood flow in the peripheral blood vessels cause fluctuations in the ppg signal in the same frequency range as the respiration signal.

11.4.3 MEMS accelerometers and barometric pressure sensors for respiration rate, heart rate and fall detection

More than 15 years ago accelerometers were added to the first smart phones from Nokia. As a result of the mass fabrication the cost of this component has been reduced by a large factor and at the same time the quality of the measurement has improved. Note that the power consumption of these devices is extremely low, in the order of 10's of μW for a 3 axis sensor with a digital output signal of 16 bits. Furthermore gyroscopes and barometric pressure sensors are now used in smart phones and sport devices. Examples of the components and medical devices that use accelerometers are shown in Figure 11-18. The area of the components is typically a few square millimeters. The data output is digital, the maximum number of bits is 16.

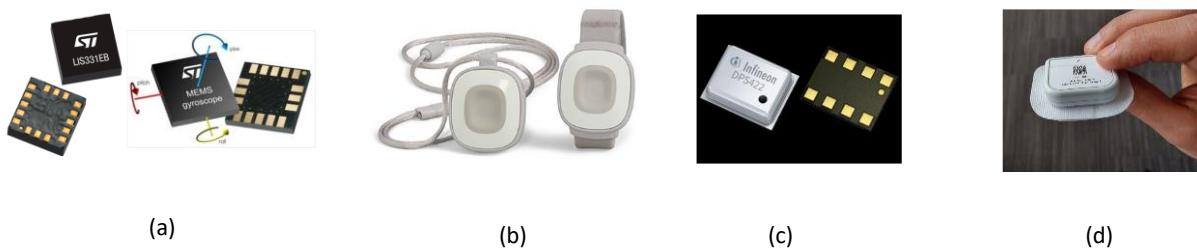


Figure 11-18 MEMS accelerometers and gyroscopes (a), fall detection devices for elderly (b), barometric pressure sensor (c) and Healthdot device (d).

These sensors are now used in many products for activity and posture measurements (Figure 11-18b), for fall detection (Figure 11-18b), for heart rate and respiratory rate in the Healthdot Figure 11-18d and for barometric based height sensing in sport watches and navigation systems and for fall detection (Figure 11-18c). Both accelerometers and barometric height are used for reliable sensing for fall detection in products of Philips lifeline in the USA (monitoring of elderly). Similar technology is now used in the latest smart watches like the Apple series 8 smartwatch. MEMS based sensing devices will be used on a large scale in modern wearable devices.

11.4.4 Blood Pressure

For invasive blood pressure measurements the errors due to height differences between measurement location and sensor position and distortion of the waveform due to second order filtering effects of the catheter-sensor system need to be reduced. A MEMS pressure sensor solution would be ideal but non-invasive blood pressure catheters are single use devices and presently the use of such devices is hindered by too high cost. Furthermore patient monitors are not equipped to handle use MEMS based pressure sensing. Finally complications such as infections and bleeding are still major issues.

There is a need for non-invasive, less obtrusive and more continuous blood pressure measurements. Furthermore the accuracy of non-invasive needs to be improved. For instance for the optimization of hypertension treatment in the home 24 hour data are required. For instance, the variations and patterns in blood pressure during day and night provide important information for the clinician.



Figure 11-19 Trends in continuous blood pressure. The invasive line method could be replaced by other techniques like volume clamp and pulse transit time.

There are several technology options for continuous non-invasive blood pressure estimates, two options are shown in Figure 11-19. The volume clamp is used in hospitals and is commercially available for many years. However, the device has not found widespread use. There are concerns of the quality and accuracy. Furthermore it is in fact still quite obtrusive, blood flow to finger tips is blocked and after a few minutes it becomes painful. This technique is not a good solution for home use.

In the consumer and home domain pulse transit time (PTT) and pulse arrival time (PAT) have been proposed for many years but despite an intense research progress is limited. There is no acceptable calibration procedure. Furthermore PTT or PAT are surrogates of an already surrogate non-invasive blood pressure measurement. Changes in PTT are often not related to blood pressure but are caused by changes in posture and bio-mechanical properties of blood vessels. Related techniques like ppg waveform analysis are also being pursued (Samsung) but are even less accurate. Note that the large tech companies are still pushing for these technologies for use in smart watches. It is not likely that these technologies will be good enough for clinical grade measurements. A similar conclusion was already mentioned for the SpO₂ measurements obtained from these devices.

For the intermittent non-invasive measurements there is also a need for less obtrusiveness (shorter measurement time, more comfort) and improvement of the measurement accuracy. The auscultatory and oscillometric measurements are widely used both in the hospital and home. Possible points of improvement are: reduction of the long measurement time, elimination of systematic errors, poor accuracy especially for hypo- and hypertension patient groups, for (morbid) obese people and neonates. Finally in the measurement data there is much more information that is presently hardly used.

The following causes of the poor accuracy of cuff based non-invasive blood pressure techniques have been mentioned: impact of cuff design and materials, the frequency dependence and magnitude of

the cuff transfer function are not well understood, the impact of arm tissue compression especially for obese persons needs to be quantified, too long measurement times lead to blood pooling in the lower arm, the impact of distal arm blood pressure(s) on the cuff-pressure signal and auscultatory signals is not well understood, the mechanical properties of the brachial artery. All effects influence the shape of the cuff pressure envelope and affect the algorithms used for estimation of blood pressure.

In summary there is no good solution for a continuous non-invasive, unobtrusive and clinical grade non-invasive blood pressure measurement. Furthermore the most widely used techniques for intermittent non-invasive blood pressure measurement need to be improved.

11.4.5 Cardiac Output

Some argue that the gold standard measurement for cardiac output (i.e the thermal dilution method) is not a real gold standard, there are doubts on its accuracy. The thermodilution technique is intermittent, highly invasive, poses high risk for the patient and depends on the time-phase in the respiratory cycle. There are several options available for continuous beat to beat stroke volume and cardiac output. These methods (like pulse-contour analysis) are model based and the parameters need to be determined by calibration with an intermittent gold standard technique. There are several drawbacks: the first is that invasive artery and venous access is needed. Furthermore techniques are model based, need frequent (re-)calibration due to changes in patient physiology, are valid only for a regular heartbeat and there are issues when centralization of blood flow occurs, i.e. when signals from peripheral blood vessels become less reliable. In short, these methods are less reliable when needed most.

During critical care, measurement of cardiac output is one of the most important measurements. Beat to beat flow information on blood flow is extreme importance for clinicians to optimize therapy. In short, there is a need to include non-invasive flow based measurements and parameters in patient monitoring devices. Due to progress in measurement technology, artificial intelligence and data analysis techniques other technologies may become available. Developments in ultra sound (US) technology can be exploited. Smart US devices enable use by non-experts, image quality is enhanced by artificial intelligence and machine learning and (local) flow information can be obtained beat-to-beat with relatively low cost solutions. Although the absolute value of flow rates may be less accurate changes in flow and its waveform contain important information. An example is the use of a patch to measure flow in the carotid artery (see Figure 11-20). Flow to the brain is not reduced during critical situations and is roughly proportional to cardiac output. It is likely that such US technologies may be included in patient monitoring devices in the near future.

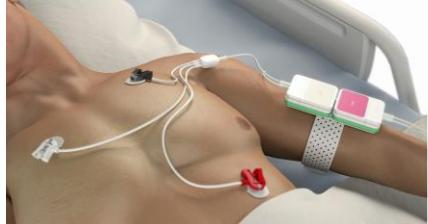


Figure 11-20 Impression of the use of carotid ultra sound patches during surgery.

11.5 Next generation Patient Monitor Devices

As discussed earlier in this chapter patient monitor architectures have remained the more or less same over the last 20 to 30 years. This is partially due to the success of these devices and its backwards compatibility (i.e. enables reuse of modules, sensors and cables, user interface). Furthermore people are used to work with these devices. But the drawbacks are becoming more and more clear especially for the new use cases. Customer interviews have shown that users want a more modern architecture, more in line with the systems like smart consumer devices. A list of the main wishes of clinicians and an artist impression of a possible solution is shown in the table below.

A possible implementation of these wishes is shown in an artist impression shown in Figure 11-21. Note that the measurements have to be of professional clinical grade, this limits miniaturization to a certain degree. Furthermore regulatory requirements on the electrical isolation of measurements, data and mains power isolation limit reduction in size. Although it is technically possible to integrate the main measurements on a single integrated circuit the high voltage isolation requirements prohibit the use of such an solution. An ASIC per measurement is possible. Note however that protection and power generation in the module will require a lot of space, often much more than the area of the measurement itself. The same holds for the connectors, they cannot be made arbitrary small. The isolation requirements can be resolved by the use of dielectric isolation between the measurement modules, the use of contactless power and wireless data transfer. A complete measurement with a newly developed connector can be of the size of a credit card with a thickness of a few mm (i.e. slightly more than the thickness of a typical Lithium Polymer battery of the required energy capacity). A battery run time for continuous use of more than 3 days is possible.

For all use cases from low- to high-acuity – a system that can travel with the patient	
<ul style="list-style-type: none"> ▪ No monitoring gaps, continuous data, efficient, connected, improves patient care and safety ▪ One scalable and modular system (from low to high acuity) ▪ Patient worn (battery life) and bedside use ▪ Less cables ▪ Backwards compatible with installed base ▪ Fits in existing workflow, reduces workload ▪ Up-to-date connectivity, networks to back-end, hospital information systems and the cloud ▪ Compatible with smart device use ▪ Guaranteed quality of service ▪ Smart alarms, warnings, predictive alerts (data) ▪ Compliant with new business models , value-based solution, pay per use 	 <p>Figure 11-21 Artist impression of a new generation of patient worn monitor devices.</p>

Backwards compatibility with existing monitor devices is very important. They can serve as the host devices for the new measurements, furthermore the complex alarm generation can still be done by the monitor device, users can reuse their existing equipment, and other invasive measurements like blood pressure can still be used using the same hardware.

New functions are possible, data from mobile patients from the home to hospital and home are available using a single system, data can be transferred to servers in the hospital or in the cloud. At these locations much more computing power and data storage capacity is available. Via the remote

servers data can be shown in real time on a smart device. What remains are the challenges with alarm fatigue and the underutilization of data and information. This will be discussed in following section of this chapter.

The solution for patient worn monitoring compatible with high acuity monitoring is not well suited for longer duration home monitoring for chronically ill patients or after discharge from hospital. This requires a dedicated device of small size, disposable, with lower cost and much longer battery run time. This was discussed in section 10.3.1.

Appendices

12 Appendices for Chapter 3

12.1 Volume-Pressure relation and Compliance for an elastic cylindrical tube

By combining the equations of Laplace and Hooke a first order volume-pressure (V-P) relation of the thin walled cylindrical element can be derived. Thereafter an expression for the differential compliance C of thin walled cylinder is derived. This is a very important mechanical and physiological parameter of elastic tubes in the circulation and respiration systems.

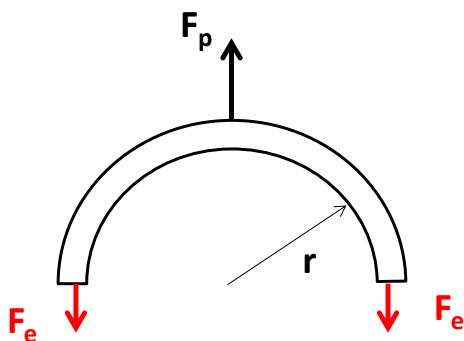


Figure 12-1 Upper half of cylindrical tube with elastic and pressure forces

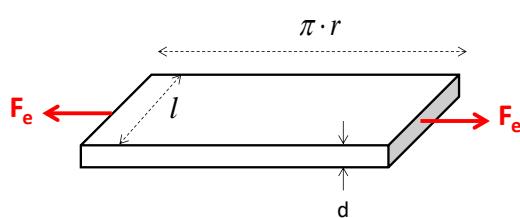


Figure 12-2 Upper half cylinder wall laid out in a plane, forces and dimensions are indicated in the figure.

From the law of Laplace, the relation between radius and transmural pressure P is obtained. For zero transmural pressure the cylinder wall is not stretched, hence the wall stress is equal to zero, the radius is r_0 and the length of the cylinder half arc is πr_0 . When the transmural pressure is increased wall stress and radius increase. The length of the half arc is now πr (see Figure 12-2). The circumferential wall strain and wall stress (assume Hooke's equation is valid) are equal to:

	$\varepsilon_w = \frac{\pi r - \pi r_0}{\pi r_0} = \frac{r - r_0}{r_0}$	12.1-1
	$\sigma_w = E \cdot \frac{r - r_0}{r_0}$	12.1-2

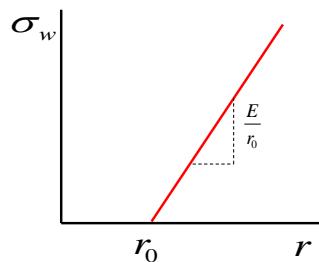


Figure 12-3: Wall stress as function of the cylinder radius (equation 12.1-2).

Equation 12.1-2 is plotted in Figure 12-3. It is important to note that wall stress is built up only when the tube radius r is larger than r_0 , i.e. when the tube volume V is larger than the unstressed tube volume V_0 . The unstressed volume is an important parameter of an elastic tube or artery. In our example a tube has an unstressed volume of $V_0 = \pi r_0^2 l$. Now a fluid (e.g. water, blood) is pumped into the tube, as long as the volume of fluid is less than V_0 no pressure is built up, when the fluid volume is larger than the unstressed volume an over-pressure is created and transmural pressure

and wall stress increase. The important observation is that a fluid volume larger than the unstressed volume is needed for a positive transmural pressure. This is very important for blood pressure in the circulation. The slope of the wall stress versus radius is E/r_0 . The stiffer the tube and the smaller the unstressed diameter the larger the wall stress for a fixed radius.

By combining equations 3.2-8 (Laplace) and 12.1-2 (Hooke) a relation between tube radius and transmural pressure can be derived (equation 12.1-3). The volume of the cylinder as function of transmural pressure P is $V = \pi r^2 l$, the volume is given by equation 12.1-4.

	$r(P) = \frac{r_0}{1 - \frac{Pr_0}{Ed}}$	12.1-3
	$V(P) = \frac{V_0}{\left(1 - \frac{Pr_0}{Ed}\right)^2}$	12.1-4

The volume of a tube with dimensions and E modulus like a human aorta are plotted in *Figure 12-4*. It will be shown later in the lecture on blood pressure measurement that equation 12.1-4 does not fit the dependence of aortic volume on transmural pressure for larger pressures, there are large deviations due to the non-homogeneous and non-linear properties of the vessel wall material.

The dependence of tube volume on transmural pressure is an important measure of the elasticity of the tube. Tube elasticity or its inverse the compliance C is a very important parameter for fluid dynamics and the cardiovascular and respiratory physiology. The impact of a volume change on transmural pressure is illustrated for an elastic tube in *Figure 12-5* for two cases.

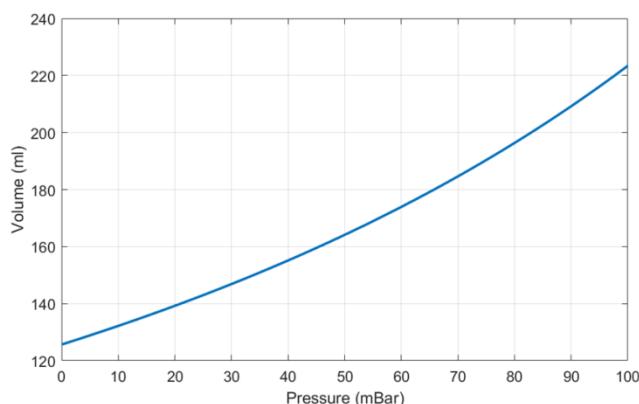


Figure 12-4 Calculated tube volume. Parameters used are $r_0: 0.01 \text{ m}$, $E: 200000 \text{ N/m}^2$, wall thickness 2mm, length 0.4 m.

In *Figure 12-5* a small volume of 20 ml is added to the tube at a filling status of 140 ml and 200 ml. The transmural pressure changes because of the increase in radius and wall stress. The pressure change is a lot larger at 140 ml than at 200 ml starting volume (note that the elastic modulus was not strain dependent). This is caused by the typical non-linear dependence of the tube volume on pressure even for the simple elastic tube that was studied here. The ratio of volume change and corresponding pressure change is a measure of the tube elasticity, the ability to stretch when a volume is added to the tube. For a stiff tube the pressure change is much larger than for a very compliant tube. The parameter that relates a volume change to a pressure change of a hollow elastic object is the compliance C. Compliance is the ability of a hollow object to distend in volume when the

transmural pressure is raised. It can be defined in two ways (see *Figure 12-6*). The differential compliance is defined by:

$$C(P) = \left(\frac{dV}{dP} \right)_{P_1} \quad 12.1-5$$

It is the tangent of the volume-pressure curve at volume V_1 and pressure P_1 . In physiology this parameter is called dynamic or differential compliance. This is the most relevant elastic parameter in physiology. In physiology in some cases the static or large signal compliance C_s is used, it is defined by:

$$C_s = \frac{V_1}{P_1} \quad 12.1-6$$

As can be seen in *Figure 12-6* the numerical values of the static and dynamic compliance differ strongly for typical biologic non-linear tissues. When the volume-pressure relation is linear, the static and dynamic compliance are identical.

The compliance (equation 12.1-7) and a first order equation for the compliance for very small strain values (equation 12.1-8) can be derived by differentiating the volume-pressure equation 12.1-4:

$$C = \frac{dV}{dP} = 2V_0 \cdot \frac{r_0}{Ed} \cdot \frac{1}{\left(1 - \frac{Pr_0}{Ed}\right)^3} \quad 12.1-7$$

$$C = \left(\frac{dV}{dP} \right)_{P=0} = 2V_0 \cdot \frac{r_0}{Ed} \quad 12.1-8$$

$$C = \left(\frac{dV}{dP} \right)_{P=0} = 2V_0 \cdot (1 - \mu^2) \cdot \frac{r_0}{Ed} \quad 12.1-9$$

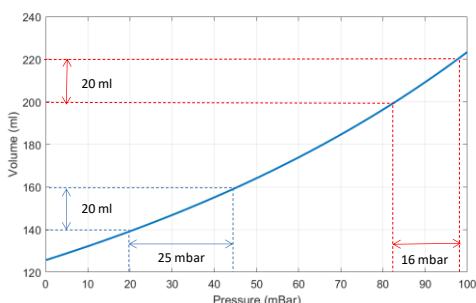


Figure 12-5 Tube volume versus pressure (same curve as in *Figure 12-4*). The addition of a 20 ml test volume and the resulting pressure change is indicated in the figure for two different volumes.

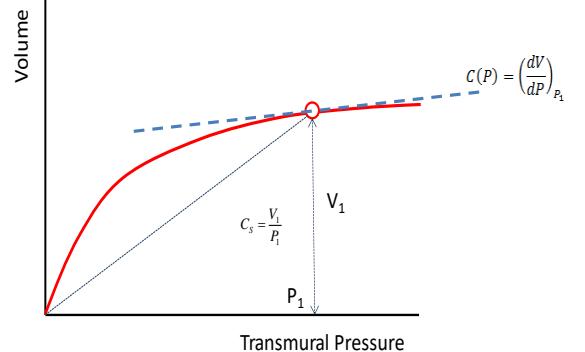


Figure 12-6: Arterial volume-pressure curve and illustration of the definition of dynamic and static compliance.

The unstressed volume is equal to $V_0 = \pi r_0^2 l$. For small pressure and small strain, the compliance can be approximated by equation 12.1-8. This equation is often cited in literature as the equation for the compliance of an ideal elastic tube at arbitrary pressure. The dynamic compliance (equation 12.1-8) is proportional to the product of unstressed volume and radius V_0 and r_0 and inversely proportional to the product of the elastic modulus and wall thickness. The ratio r_0/d hardly varies in the arterial tree. In first order for human physiology the vessel compliance is determined by the unstressed volume and the elastic modulus of the wall material. Hence smaller diameter tubes have

a smaller compliance, are stiffer. In the derivation of the volume-pressure equation the transverse contraction strain and Poisson ratio were not considered. A more rigorous derivation of the equation includes the transverse contraction and a constrained tube length l as would be the case in a realistic system. This results in the extra term $(1-\mu^2)$ in equation 12.1-9.

12.2 First and second order models of viscoelastic materials (optional)

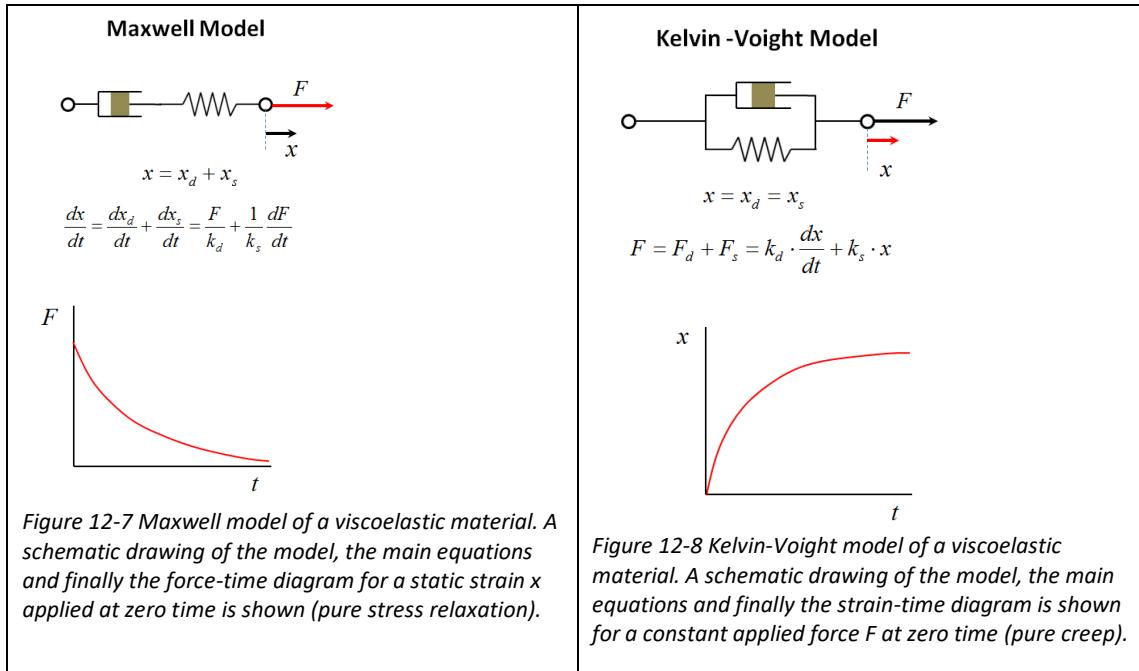
Three simple lumped element models for viscoelastic materials are discussed to illustrate pure stress relaxation (Maxwell), pure creep (Kelvin-Voight) and finally the simplest model of a viscoelastic material which includes creep and stress relaxation, the linear solid model. This last model is the simplest model that can be used to model biologic materials.

In the Maxwell model a material is modeled by an ideal dashpot and spring connected in series (see *Figure 12-7*). In a series connection the force F is equal for both elements, the total strain x is the sum of the strains of both elements. When a force is applied the strain of the elastic element increases instantaneously, the strain of the dashpot increases linearly with time as long as the force is applied (see *Figure 3-16*). When a fixed strain is applied the elastic strain is initially equal to the applied strain, the strain of the dashpot is initially zero and increases with time. Since total strain is constant and the strain of the dashpot increases, the strain of the elastic element decreases. The result is that the force decreases, the Maxwell model predicts that the force decreases exponentially with time when the strain is kept constant. Many viscoelastic materials show such stress relaxation behavior. However the creep (i.e. changes in strain after stresses are applied) is not modeled well.

The Kelvin-Voight model (see *Figure 12-8*) was proposed to model the creep of polymer materials. The model consists of a parallel spring-dashpot circuit. For this model the strains of both elements are equal, the total force is the sum of the forces of the two elements. The force-strain relation is described by a linear first order differential equation. When a constant force is applied the dashpot limits the strain rate, when time progresses the strain of the dashpot increases, and the spring is stretched. The strain rate decreases exponentially towards zero. The spring dominates the strain at larger times, the limiting value is of the strain is F/k_s . This model describes creep of solid viscoelastic materials well but does not include stress relaxation. For realistic materials a more complex model with a connection of multiple serial and parallel spring-dashpot combinations is needed.

The simplest model for a viscoelastic solid is the standard linear solid model. It is shown in *Figure 12-9*. This model combines a Maxwell model with a linear spring connected in parallel. The initial strain x_0 is purely elastic (parallel combination of spring 1 and 2). Thereafter it deforms further towards an asymptotic strain value, this is the viscous component of the strain. The asymptotic value is equal to the strain of spring 1 due to the force F . This model is more accurate than the previous models the description of the stress-strain relation, its time dependence does not describe measurements well. In practice more complicated series parallel combinations are needed since deformation is not governed by a single time constant as is implicitly assumed in this model.

In previous sections the mechanical behavior of elastic and viscoelastic materials was discussed. The mass of the material was not included in the models. This is a very important parameter, inclusion of mass and inertia in the models is discussed in the next section.



Standard Linear Solid Model

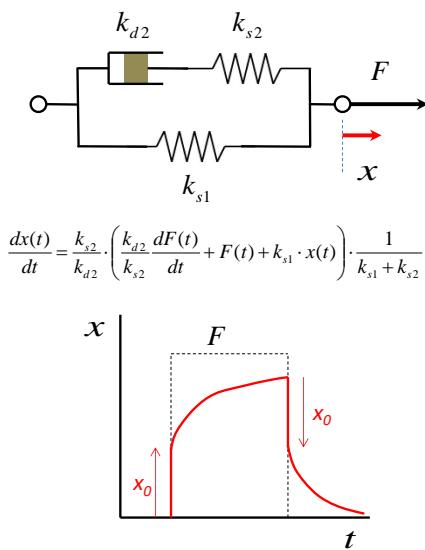


Figure 12-9 Schematic drawing of the schematics of the standard linear solid model, the governing differential equation that relates stress and strain and an illustration of the time dependence of the strain for a force pulse F .

12.3 Damped spring-damper system with Inertia and Mass

An object with mass m is connected to a spring and damper, the gravitational force stretches the spring and the mass reaches a rest position x_0 . When the mass is pulled from its rest position and is released it will oscillate about its rest position, the oscillation amplitude will dampen out and finally the mass will come to rest. This cannot be described by the models used in the previous section. The oscillation is caused by the mass of the object and its inertia. Inertia is the resistance of a physical object to change its motion, this includes changes in velocity, direction and state of rest. Inertia is one of the fundamental principles in classical physics that describes changes in motion due to applied forces. Inertia is proportional to the mass of an object. Newton's first law states that when the net

force on an object is zero, the acceleration is zero and the object moves with a constant velocity. A special case is when the velocity is zero, the object is in rest. An inertial reference system is a coordinate system with no acceleration of the reference frame. In the following sections the objects are in an inertial reference system with reference velocity equal to zero.

The velocity or direction of motion of an object changes when a net force is exerted on the object. This is formulated in Newton's second law which states that the rate of change of the velocity of an object is proportional to the net force exerted on the object and inversely proportional to the mass of the object. In the form of an equation:

$$\sum_{n=1}^k F_n = m \cdot \frac{dv}{dt} = m \cdot a \quad 12.3-1$$

Both the force and acceleration are vectors, in the remainder of the text only one-dimensional systems are discussed for illustration purposes. In this case the sign of the force is important, it is linked to the direction of the force. The sign is positive when the force is in the positive coordinate direction and negative when it is in the opposite direction.

12.3.1 Dynamic models of viscoelastic inertial systems

A first order model is used to illustrate system behavior and time dependence of motion for an elastic model system that includes inertance, elasticity and damping. In *Figure 12-10* a piece of material is driven with an external force F_{ext} and has mass m , length l_0 and area A , Young Modulus E and damping coefficient η_d . This model is complex and often simpler approximate models are used that still contain the essence of the full model. In our example the dynamic deformation of the object is approximated by a well-known lumped element model. An equivalent lumped element model of the object consists of a point mass m , a damper and an elastic spring, it is shown in the right part of this figure. The equations of motion of the lumped model are easier to solve, the system behavior is similar.

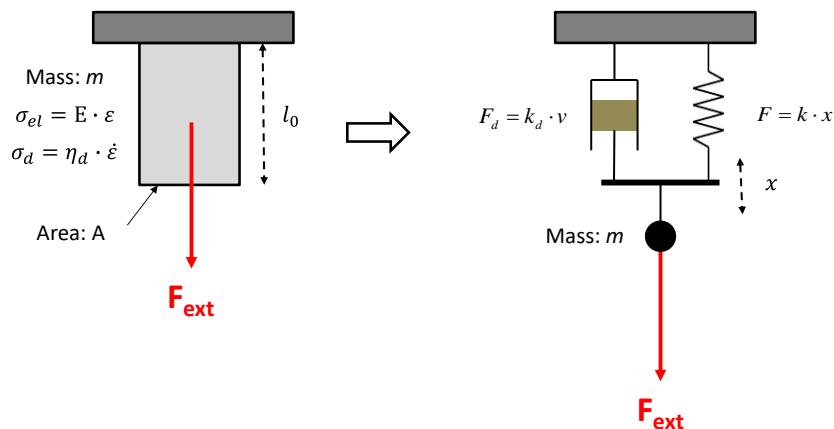


Figure 12-10 Continuous viscoelastic material (left) with length l_0 , area A and mass m and equivalent lumped element model that consists of a spring-damper-mass system with deviation x from the reference position (right).

The equation of motion for the lumped model follows from the second law of Newton (equation 12.3-1). The equation of motion (eq. 12.3-2) is given below.

$$\sum_n F_n = F_{ext} - k \cdot x - k_d \cdot \frac{dx}{dt} = m \cdot \frac{d^2x}{dt^2} \quad 12.3-2$$

$$F_{ext}(t) = m \cdot \frac{d^2x}{dt^2} + k_d \cdot \frac{dx}{dt} + k \cdot x \quad 12.3-3$$

$$\frac{d^2x}{dt^2} + 2\zeta\omega_0 \frac{dx}{dt} + \omega_0^2 x = 0$$

12.3-4

The damping parameter ζ and the undamped resonance frequency ω_0 are the two important parameters that determine the frequency response of the system.

$$\zeta = \frac{k_d}{2\sqrt{m \cdot k}}, \omega_0 = \sqrt{\frac{k}{m}}$$

Rearranging terms in equation 12.3-2 gives equation 12.3-3 which is a homogeneous linear time invariant second order differential equation. Equation 12.3-4 can be solved exactly for any driving force $F_{ext}(t)$ when the parameters ζ and ω_0 are known. These parameters depend on material properties and the mass of the system. The solution depends on the external force, value of ζ and boundary conditions. Solutions for different values of ξ are:

- $\zeta > 1$: overdamped system, the system returns to the steady state without oscillations (only exponential decay), the larger the value of ξ the longer it takes to reach the steady state condition.
- $\zeta = 1$: critically damped system, the system returns to steady state as quickly as possible without oscillations.
- $0 < \zeta < 1$: underdamped system, the system oscillates with a frequency that depends on both ζ and ω_0 . The system returns with an exponentially decreasing oscillation amplitude to steady state. The smaller is ζ the longer it takes to reach steady state.
- $\zeta=0$: undamped system, the system oscillates with constant amplitude and angular frequency ω_0 . This is the ideal harmonic oscillator.

For ζ larger than zero energy is dissipated in the system in the form of heat that is transferred from the object to the environment, hence energy is lost. This loss causes an exponential relaxation of the motion towards a steady state condition. For $\zeta=0$ no energy is dissipated in the system and after excitation, elastic and kinetic energy is conserved, the system oscillates forever, elastic energy is converted in kinetic energy and vice versa.

We will consider two types of external driving force, the first one is the step force (force equal to zero for time less than zero and constant F_0 for other times) and the second is a harmonic sinusoidal excitation (force with amplitude F_0 and angular frequency ω).

12.3.2 System with a step function driving force

For this type of external driving force the transient solution for $x(t)$ and $\zeta < 1$, $x(0) = 0$, $F_0 = k$ (i.e. force for unit displacement of x) is given in equations 12.3-5 and 12.3-6.

$$x(t) = 1 - e^{-\zeta\omega_0 t} \frac{\sin(\sqrt{(1-\zeta^2)}\omega_0 t + \varphi)}{\sin \varphi}$$

12.3-5

$$\cos \varphi = \zeta$$

12.3-6

The oscillation frequency is $\sqrt{(1-\zeta^2)}\omega_0$ and the exponential decay factor is $\zeta\omega_0$. Solutions for equation 2-28 are plotted in for various values of ξ between 0 and 1 in *Figure 12-11*. The time $\tau = 1/(\zeta\omega_0)$ is the relaxation time, it is a measure of the time that the system needs to adapt to changes in driving forces.

The displacement depends on the parameter ζ , for the undamped case the oscillation is permanent with the natural oscillation frequency ω_0 . In this case no energy is dissipated and motion energy and elastic energy are conserved. When there is damping in the system the oscillations dampen out exponentially with a time constant τ due to energy dissipation. The larger ζ the faster oscillations damp out. For $\zeta \geq 1$ there are no oscillations possible anymore.

In engineering oscillatory behavior can be used or must be limited. For instance, for a pendulum clock or a tuning-fork the damping parameter ζ must be as small as possible. For other systems such as the wheel-suspension in a car, the control of the wheel oscillations due to bumps in the road is crucial, too much and too little damping is not good. In this case the system is designed such that the damping parameter ζ is in the range of 0.2-0.3 to one. In many cases the driving force is not a step but a periodic harmonic force. In the following section the system behavior when a harmonic driving force is applied is shown.

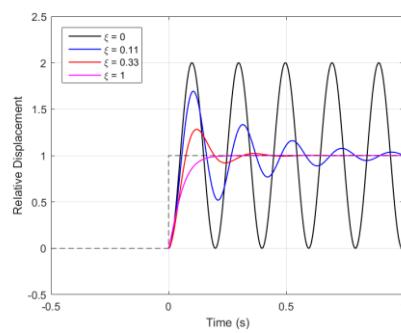


Figure 12-11 Relative displacement normalized on the static displacement versus time for a spring-damper-mass lumped element model with spring constant k of 1000 N/m, mass 1 kg, F_0 1000N and damping ratios of 0, 0.11, 0.33 and 1. The dashed line represents the step function.

12.3.3 System with harmonic driving force

As an example, a harmonic sinusoidal driving force with amplitude F_0 and angular frequency ω is applied to the spring-damper-mass system with parameters m , ω_0 and ζ . The equation of motion is given equation 12.3-7. The solution for the steady state condition (typically reached after a few oscillations) is given in equations 12.3-8 and 12.3-9 as function of the angular frequency (ω_n is the ratio of ω and ω_0). The phase angle φ is given in equation 12.3-10.

	$\frac{d^2x}{dt^2} + 2\zeta\omega_0 \frac{dx}{dt} + \omega_0^2 x = \frac{F_0}{m} \sin(\omega t)$	12.3-7
--	--	--------

	$x(t) = A \cdot \sin(\omega t + \varphi)$	12.3-8
--	---	--------

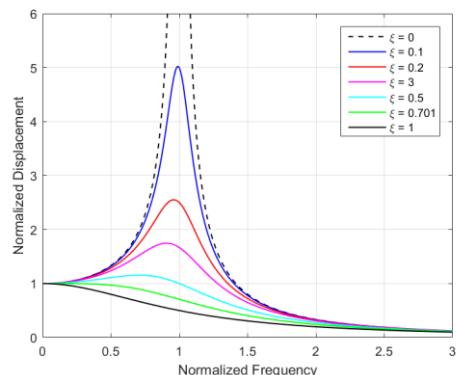
	$A = \frac{F_0}{k} \cdot \frac{1}{\sqrt{(1 - \omega_n^2)^2 + (2\zeta\omega_n)^2}}$	12.3-9
--	--	--------

	$\varphi = \tan^{-1} \left(\frac{2\zeta\omega_n}{1 - \omega_n^2} \right)$	12.3-10
--	--	---------

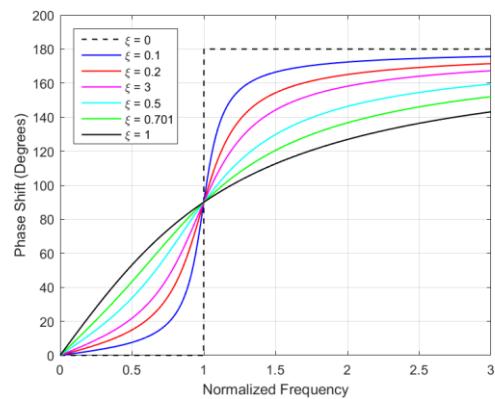
The solution for displacement is also a harmonic function, the amplitude A and phase angle φ depend on the parameters F_0 , k and ζ and the normalized frequency. In figure 12-12 and figure 12-13

the normalized vibration amplitude and phase angle are plotted versus normalized frequency ω_n . These plots show several important features.

With increasing frequency, the amplitude reaches a maximum and then decreases. For a specific frequency called resonance frequency the displacement is largest, the resonance frequency is equal to $\sqrt{(1 - \zeta^2)}\omega_0$. Resonance can only occur for $\zeta < 1/\sqrt{2}$. When the driving frequency is near the resonance frequency of the object, very large vibration amplitudes can be generated and consequently a large amount of energy is transferred into the system. For low frequency (i.e. much smaller than the resonance frequency) the displacement is proportional to the force and is close to the static displacement of the spring. The mass and damper hardly influence displacement. The driving force and displacement are in phase. For very large frequency (frequency much larger than the resonance frequency) the displacement is very small and the phase angle approached 180 degrees, i.e. driving force and displacement are maximum out of phase. Here the inertance dominates motion, the object cannot follow the rapid change in force anymore. For small ζ and near resonance frequency the vibration amplitude becomes very large and the phase angle is 90 degrees (i.e. the displacement lags the force with 90 degrees). In this case the driving force is always in the direction of motion. For each period the driving force pushes the mass in the same direction, i.e. the amplitude and velocity increase with each cycle. The kinetic and elastic energy increase at each cycle. When damping is present part of the energy is converted into heat that is transferred to the environment. Hence damping limit displacements and the total energy that is stored in the system. With increasing amplitude, the velocity of the object increases and friction will oppose motion more and more. When the friction is large, oscillation is not possible anymore



12-12 Frequency response of the amplitude of a damped spring-mass system versus normalized frequency ω_n for various values of ζ . The displacement is normalized to the static displacement. ($k: 1000 \text{ N/m}$, $m=1 \text{ kg}$)



12-13 Frequency response of the phase angle of a damped spring-mass system versus normalized frequency ω_n for various values of ζ . The displacement is normalized to the static displacement. ($k: 1000 \text{ N/m}$, $m=1 \text{ kg}$)

In reality the driving force is not harmonic but the harmonic model can still be used by decomposing the repetitive driving force into its harmonic Fourier components and performing the analysis for each harmonic and adding up the displacements of the harmonics.

13 Appendices for Chapter 4

13.1 Navier Stokes Equation

Conservation laws and Navier-Stokes Equations

For a closed system the following conservation laws hold:

1. Conservation of linear momentum p . The total momentum $\vec{p} = m \cdot \vec{V}$ is constant, m is the total mass and \vec{V} is the velocity field.
2. Conservation of energy E . The total energy in the system cannot change, one form of energy can be converted to another form but the sum is constant (e.g. potential energy to kinetic energy etc.).
3. Conservation of mass. The mass of a closed system remains constant.

The equations of motion of classical mechanics are used, i.e. the three equations from Newton:

1. When the net force on an object is zero the object is either at rest or moves with a constant velocity ($\sum F = 0$).
2. The vector sum of forces is equal to the product of mass and acceleration ($\sum \vec{F} = m \cdot \vec{a}$)
3. When one body exerts a force on a second body, the second body simultaneously exerts a force equal in magnitude and opposite in direction on the first body (action = reaction).

The Navier-Stokes equation is derived by applying the above conservation laws and Newton's laws on an infinitesimal small element with volume $dxdydz$ at location xyz in the system (i.e. a small control volume). In this lecture only three types of force are taken into account, the driving force due to pressure gradients, friction and gravity. In this course the formulas for cylindrical tubes are derived. Therefore the NS equation is formulated for cylinder coordinates (see Figure 13-1).

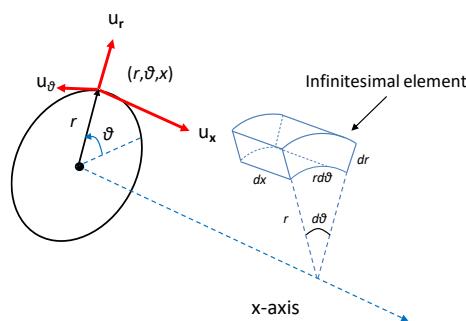


Figure 13-1 Cylinder coordinate system. The coordinates and velocity components u are shown in the figure.

For a tube filled with an incompressible Newtonian fluid the NS equation is:

	$\rho \frac{\partial \vec{V}}{\partial t} + \rho (\vec{V} \cdot \nabla) \vec{V} = -\nabla P + \eta \Delta \vec{V} + \rho \vec{g}$	13.1-1
--	---	--------

13.1-1

Here $\vec{V} = (u_r, u_\theta, u_x)$ is the velocity vector at coordinates (r, θ, x) , \vec{g} is the gravitation vector at these coordinates, ∇ is the nabla operator and Δ is the Laplace operator. The second term on the left ($\rho(\vec{V} \cdot \nabla)\vec{V}$) is the convective acceleration term that is needed for pipe geometries with changing diameter or shape. For straight pipes this term can be neglected. For systems where viscosity forces can be neglected the NS equation simplifies to the Euler equation:

$$\rho \frac{\partial \vec{V}}{\partial t} + \rho (\vec{V} \cdot \nabla) \vec{V} = -\nabla P + \rho \vec{g}$$

13.1-2

Finally the equation of conservation of mass for incompressible fluids is:

$$\frac{1}{r} \frac{\partial(r u_r)}{\partial r} + \frac{1}{r} \frac{\partial(u_\theta)}{\partial \theta} + \frac{\partial u_x}{\partial x} = 0$$

13.1-3

13.2 Basic Network solution methods

13.2.1 Network Relations

The following relations and equations are essential for modeling of tube flow in networks of tubes and active elements.

13.2.2 Kirchhoff relations

The Kirchhoff current law (KCL first term in equation 13.2-1) states that the sum of the k currents that flow into and out of a node must be equal to zero (see Figure 13-2). The sign of the current is positive when the current flows into the node and negative when the current flows away from the node.

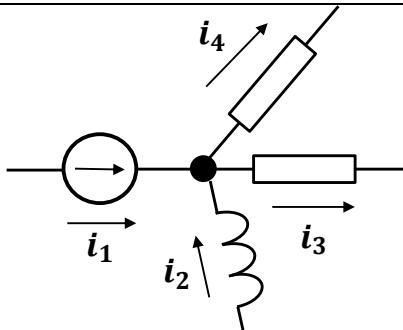


Figure 13-2 Kirchhoff current relations.

$$i_1 + i_2 - i_3 - i_4 = 0$$

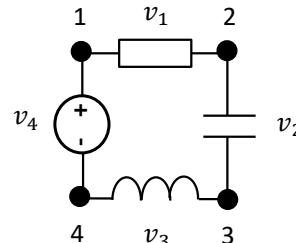


Figure 13-3 Kirchhoff voltage law.

$$v_{12} + v_{23} + v_{34} + v_{41} = 0$$

$$\sum_{n=1}^k i_n = 0 , \sum_{n=1}^k v_{n,n+1} = 0$$

13.2-1

The sum of the currents for this specific example is shown below the Figure 13-2. This law is based on the conservation of charge. In the fluid domain this equation holds for the sum of volumetric flows towards a node (conservation of volume). The Kirchhoff voltage law (KVL second term equation 13.2-1, see Figure 13-3) states that the directed sum of all voltage differences over the k elements in a closed network is zero (conservation of energy). It is important to consistent in the direction of the indices in the loop as $v_{12} = -v_{21}$. The same equation holds for pressure differences in the fluid domain.

	$\frac{1}{R} = \sum_{i=1}^n \frac{1}{R_i}$
	$R = \sum_{i=1}^n R_i$
Voltage drop over resistor R_i in series combination	$V_i = \frac{R_i}{R} V$

Using both Kirchhoff laws relations for the replacement resistance R for a series and parallel combination of resistors and inductors can be derived. For the series combination R is equal to the sum of all resistors R_i . For the parallel combination R can be determined from the inverse sum of the resistors. The same formulas hold for inductances. For capacitors the replacement capacitance C is derived from charge relations. The results for parallel and series combination are shown below.

	$C = \sum_{i=1}^n C_i$
	$\frac{1}{C} = \sum_{i=1}^n \frac{1}{C_i}$
Voltage drop over capacitor C_1 in series	$\frac{V_1}{V} = \frac{C_2}{C_1 + C_2}$

Note that for two capacitors in series the largest voltage drop is over the smallest capacitance.

13.2.3 Recap Basic Circuits

In this section one simple RC circuit is solved as example. The results for a three element RLC are show because they are very important for both sensors and vessel segments. In appendix B more examples and the derivation of the results is described. The circuits are relevant for the simplest physiological models of the respiration and circulation systems.

13.2.3.1 Series RC circuit drawings

This simple circuit consists of a resistor and capacitor in series (see Figure 13-4). This circuit is relevant for the respiration system. Three different solution methods and the circuit properties are shown for illustration purposes.

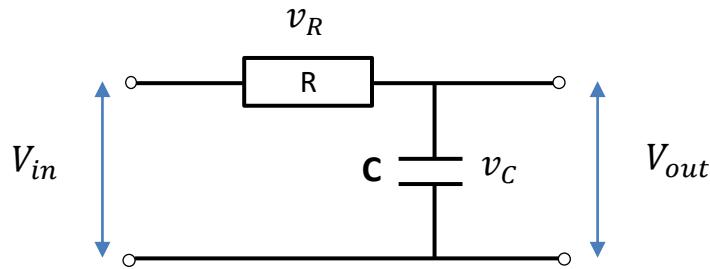


Figure 13-4 Series RC circuit.

I. Time domain

The solution is suited for transient solutions. As an example a step input voltage V_{in} that starts at $t=0$ is applied at the input terminals. We want to know the output voltage as function of time. Applying the KVL gives $V_{in} = v_R + v_C$, at $t = 0$ $v_R = v_C = 0V$. Applying KCL to the node that connects the resistor and capacitor gives $i_R = i_C$ (equation 1). Combining the KCL and KVL equations gives equation 2 for the voltage over the capacitor. Defining a new variable V^* (equation 3) gives equation 4. The solution of this differential equation (4) is an exponential function with two integration constants C_1 and C_2 . These integration constants can be determined from the boundary conditions at $t = 0$ ($V^* = -V_{in}$), $t = \infty$ ($V^* = 0$).

This gives the final solution (6) for the voltage v_c which is equal to the output voltage of this circuit. The dominant parameter of this solution is the product RC , the RC time which is the time required to reach $(1-1/e)$ of the output voltage (see Figure 13-5). It is a measure of the response time of the circuit. It takes a time of the order of the product RC before the output voltage reaches 63% of its maximum output voltage.

(1) $\frac{v_R}{R} = C \frac{dv_c}{dt}$	(2) $\frac{V_{in} - v_c}{R} = C \frac{dv_c}{dt}$
(3) $V^* = v_c - V_{in}$	(4) $\frac{dV^*}{dt} = -\frac{V^*}{RC}$
(5) $V^* = C_1 e^{-t/RC} + C_2$	(6) $v_c = V_{in} \left(1 - e^{-\frac{t}{RC}} \right)$

The response of the circuit to a step input voltage is shown in Figure 13-5 for a RC circuit with three different RC times. With increasing RC times it takes a longer time to reach the final voltage level. This RC circuit is a type of filter where low frequency components (with respect to $f_0 = 1/(2\pi RC)$ Hz) are transmitted with low attenuation and small phase lag to the output terminal. High frequency components in the signal are filtered out. The frequency response of this so-called low pass filter is studied best for pure harmonic input signals which is discussed next.

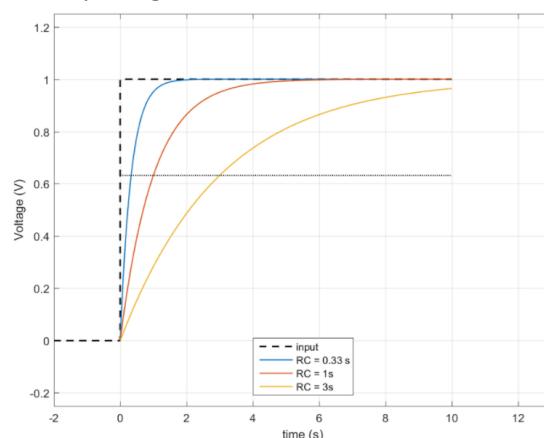


Figure 13-5 Input (step voltage of 1V) and output voltages of a series RC circuit for three RC times.

II. Complex Impedance method

A harmonic input voltage i.e. a sinewave with amplitude A and angular frequency ω is fed to the input terminals. The output signal is obtained using complex impedances Z with j the imaginary unit $j^2 = -1$. The impedances Z for a resistor, capacitor and inductor are:

$$Z_R = R$$

$$Z_C = 1/j\omega C$$

$$Z_L = j\omega L$$

The rules for parallel and series combinations of impedances are the same as for resistors. The Kirchhoff relations are valid. The advantage is that the solution is done in the algebraic domain instead of solving differential equations as was done in the previous section. For the simple RC circuit the transfer function $H(\omega)$ is the ratio between the output and input voltages (equation 13.2-2).

$$H(\omega) = \frac{V_{out}}{V_{in}} = \frac{Z_C}{Z_R + Z_C} = \frac{1/j\omega C}{R + 1/j\omega C}$$

13.2-2

$$\left| \frac{V_{out}}{V_{in}} \right| = \sqrt{\frac{1}{1 + (RC\omega)^2}}$$

13.2-3

$$\varphi = \tan^{-1} \frac{Im(H)}{Re(H)} = \tan^{-1}(-\omega RC)$$

13.2-4

The amplitude of $H(\omega)$ is given by equation 13.2-3 and the phase angle (φ) between the output and input voltages is given by equation 13.2-4. The amplitude and phase angle of the transfer function of a low pass RC circuit are shown in Figure 13-6 and Figure 13-7.

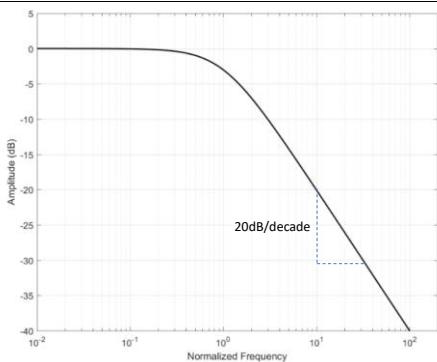


Figure 13-6 Amplitude (in dB) of the transfer function as function of the normalized frequency (normalized to $1/(2\pi RC)$).

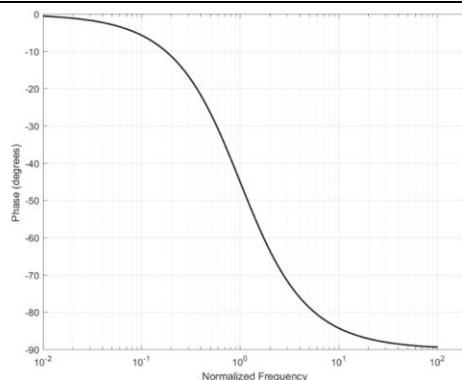


Figure 13-7 Phase angle of the output versus the input signal in degrees versus the normalized frequency (normalized to $1/(2\pi RC)$).

III. Laplace transform method RC circuit

This method is suited for both transient and harmonic analysis. A complex frequency s is defined as $s = \sigma + j\omega$. The Laplace transform of a function $f(t)$ is defined by: $F(s) = (\mathcal{L}\{f(t)\})(s) = \int_0^\infty f(t)e^{-st} dt$. The impedances in the Laplace transform method are:

$$Z_R = R$$

$$Z_C = \frac{1}{Cs}$$

$$Z_L = sL$$

The KVL (eqn 13.2-5) and Laplace transformation of the KVL equation with an input voltage equal to a step function V_{in} (eqn 13.2-6) is:

$$V_{in}(t) = v_R + v_c = i(t).R + \frac{1}{C} \int i(t)dt$$

13.2-5

	$V_{in}(s) = I(s).R + \frac{1}{Cs}I(s) = \frac{ V_{in} }{s}$	13.2-6
	$I(s) = \frac{C \cdot V_{in} }{1 + RCs}$	13.2-7

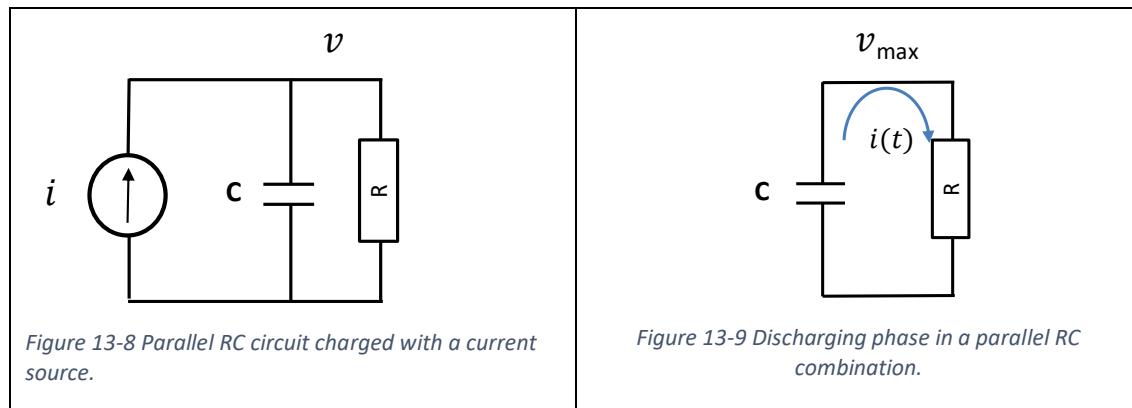
	$V_C(s) = Z_C \cdot I(s) = \left(\frac{1}{s} - \frac{1}{\frac{1}{RC} + s} \right) V_{in} $	13.2-8
--	--	--------

	$v_C(t) = V_{in} \left(1 - e^{-\frac{t}{RC}} \right)$	13.2-9
--	--	--------

From eqn 13.2-6 the Laplace transform of the current $i, I(s)$ (eqn 13.2-7) can be obtained. The Laplace transform of the voltage over the capacitor is simply the product of $Z_C \cdot I(s)$ and the inverse Laplace transformation of eqn 13.2-8 gives the voltage over the capacitor in the time domain which is equal to equation 13.2-9. The advantage of the Laplace method is that it solves the equations with algebra for any input function for V_{in} .

13.2.4 Parallel RC segment

A parallel combination of a resistor and capacitor with a current source is shown in [Figure 13-8](#). This circuit is relevant for a Windkessel model of the circulation. The voltage over the resistor and capacitor is equal. The constant current i (amplitude I_0 , step function at $t=0$) is divided over the capacitor and resistor. The voltage v at time $t = 0$ is equal to zero Volt. The Laplace method is used to derive an expression for the voltage $v(t)$. Applying the KCL for this circuit gives equation 4-113 and 4-114.



	$i = i_R + i_C$	13.2-10
	$i(t) = \frac{v(t)}{R} + C \frac{dv(t)}{dt}$	13.2-11
	$I(s) = \frac{I_0}{s} = \frac{V(s)}{R} + C(sV(s) + v(0)) = \frac{V(s)}{R} + CsV(s)$	13.2-12
	$V(s) = I_0 \frac{1/RC}{s(1/RC + s)}$	13.2-13
	$V(s) = I_0 R \left(\frac{1}{s} - \frac{1}{\frac{1}{RC} + s} \right)$	13.2-14
	$v(t) = I_0 R \left(1 - e^{-t/RC} \right)$	13.2-15

Equation 13.2-12 is the Laplace transform of equation 13.2-11. From 13.2-12 equations 13.2-13 and 13.2-14 are derived. The inverse Laplace transform of eqn 13.2-14 gives the time domain solution for $v(t)$ (eqn 13.2-15). The maximum voltage at $t = \infty$ is I_0R . This voltage varies exponentially with time, the time constant is RC .

The next example is the discharge phase of the parallel RC combination *Figure 13-9*. At time $t = 0$ the voltage is equal to V_0 . Applying the KVL to this circuit gives equation 13.2-16.

	$C \frac{dv}{dt} = \frac{v}{R}$	13.2-16
--	---------------------------------	---------

	$C(sV(s) + v(0)) = \frac{V(s)}{R}$	13.2-17
--	------------------------------------	---------

	$V(s) = V_0 \frac{1}{\frac{1}{RC} - s}$	13.2-18
--	---	---------

	$v(t) = V_0 e^{-t/RC}$	13.2-19
--	------------------------	---------

The Laplace transform of equation 13.2-16 is 13.2-17 which can be written as equation 13.2-18. The inverse Laplace transform of eqn 13.2-18 gives the voltage in the time domain (eqn 13.2-19). The voltage decreases exponentially to zero Volt with time constant RC .

13.2.5 Series RLC segment

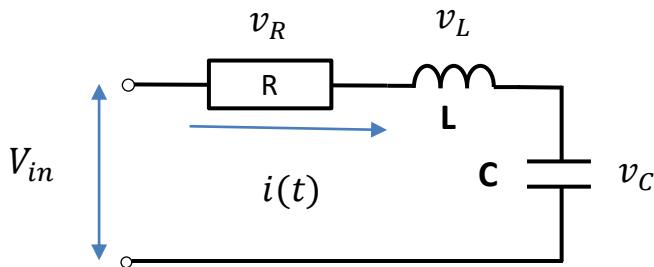


Figure 13-10 series LCR circuit with arbitrary input voltage.

This circuit is relevant for the arterial part of circulatory system and for sensor systems. The Laplace method is used to derive an expression for the transient current $i(t)$ for an arbitrary input voltage signal V_{IN} . The KVL equation (4-123) is written in terms of the current $i(t)$ in 4-124. The boundary conditions for $i(0)$ and $v_C(0)$ are both zero.

KVL	$V_{IN} = v_R + v_L + v_C$	13.2-20
KVL	$V_{IN} = iR + L \frac{di}{dt} + \frac{1}{C} \int idt$	13.2-21
LT KVL	$V(s) = RI(s) + LS(s) + \frac{1}{Cs} I(s)$	13.2-22
LT KVL	$I(s) = V(s) \frac{1}{R + LS + 1/Cs}$	13.2-23
pars	$\alpha = \frac{R}{2L}, \omega_0 = \frac{1}{\sqrt{LC}}, \gamma = \frac{\alpha}{\omega_0} = \frac{R}{2} \sqrt{\frac{C}{L}}$ $\omega_d = \omega_0 \sqrt{1 - \gamma^2}, \omega_r = \omega_0 \sqrt{\gamma^2 - 1}$	13.2-24
LPT	$I(s) = \frac{V(s)}{L} \frac{1}{s^2 + 2\alpha s + \omega_0^2} = \frac{V(s)}{L} \frac{1}{(s + \alpha)^2 + \omega_0^2 - \alpha^2}$	13.2-25

$(\gamma < 1)$	$i(t) = \frac{1}{L} \int_0^t V_{IN}(t-\tau) e^{-\alpha\tau} \left(\cos(\omega_d \tau) - \frac{\alpha}{\omega_d} \sin(\omega_d \tau) \right) d\tau$	13.2-26
$(\gamma = 1)$	$i(t) = \frac{1}{L} \int_0^t V_{IN}(t-\tau) e^{-\alpha\tau} (1 - \alpha\tau) d\tau$	13.2-27
$(\gamma > 1)$	$i(t) = \frac{1}{L} \int_0^t V_{IN}(t-\tau) e^{-\alpha\tau} \left(\cosh(\omega_r \tau) - \frac{\alpha}{\omega_r} \sinh(\omega_r \tau) \right) d\tau$	13.2-28

The Laplace transform of 13.2-21 is equation 13.2-22. A number of parameters (including resonant angular frequency ω_0 and damping parameter γ) are defined to make the equations more compact and more suitable for the inverse transformations. Then equation 13.2-23 can be written as equation 13.2-25 which can be converted to the time domain by the inverse Laplace transform of $I(s)$ with the input voltage signal. The transient solution for the underdamped case ($\gamma < 1$) is given equation 13.2-26, for the critically damped case ($\gamma = 1$) in equation 13.2-27 and for the overdamped case ($\gamma > 1$) in equation 13.2-28. Once the transient current $i(t)$ is known the voltages over the components can be calculated.

Three plots of under- ($\gamma = 0.1$), critical- ($\gamma = 1$) and over-damped ($\gamma = 4$) circuits are shown in [Figure 13-11](#). The input signal is a unit step and the output signal is the voltage over the capacitor. For the underdamped case there is significant ringing, distortion and a large overshoot of the output signal can be observed (maximum overshoot of a factor of two occurs for $\gamma = 0$) that is damped exponentially. The overshoot occurs in catheters used for invasive blood pressure measurements. This ringing and overshoot is equivalent to the behavior observed in mechanical spring-damper-mass systems, the equations are identical. The spring corresponds with the compliance or capacitor, the mass with the inertance or inductance and the viscosity with the resistor.

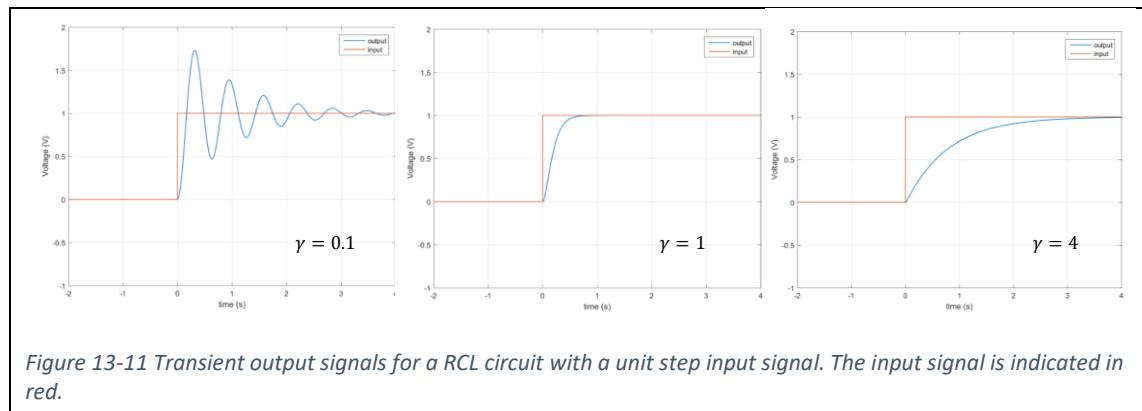


Figure 13-11 Transient output signals for a RCL circuit with a unit step input signal. The input signal is indicated in red.

Analogous to the RC circuits this circuit can also be seen as a filter where the amplitude and phase delay depend on the frequency of the input signal. The filter function can be best studied using the steady state harmonic analysis method, i.e. using pure continuous sine waves as input signal and study the relation (i.e. transfer function) with the output voltage amplitude and phase. The transfer function $H(\omega)$ can be calculated in the Laplace domain by replacing the complex frequency s with $j\omega$. $H(\omega)$ is the ratio of the output voltage over the capacitor to the input voltage. It can be derived in the Laplace domain as follows. The transfer function $H(s)$ is simply the ratio of the impedance Z_C over the total impedance (13.2-29). Replacing s by $j\omega$ and using the parameters ω_0 and γ gives the expression for transfer function $H(\omega)$ (13.2-30). The magnitude of $H(\omega)$ is the root of the product of the original and complex conjugate of $H(\omega)$ (eqn 13.2-31). The phase is given by 13.2-32.

	$H(s) = \frac{1/Cs}{R + Ls + 1/Cs} = \frac{\omega_0^2}{(s + \alpha)^2 + \omega_0^2 - \alpha^2}$	13.2-29
	$H(\omega) = \frac{1}{\left(1 - \frac{\omega^2}{\omega_0^2}\right) + 2j\gamma \frac{\omega}{\omega_0}}$	13.2-30
	$ H(\omega) = \sqrt{\left(1 - \frac{\omega^2}{\omega_0^2}\right)^2 + 4\gamma^2 \frac{\omega^2}{\omega_0^2}}$	13.2-31
	$\varphi = \tan^{-1} \frac{2\gamma \frac{\omega}{\omega_0}}{\left(1 - \frac{\omega^2}{\omega_0^2}\right)}$	13.2-32

These relations are dominated by two parameters, the resonant angular frequency ω_0 ($= \frac{1}{\sqrt{LC}}$) and the dimensionless damping parameter γ ($= \frac{R}{2} \sqrt{\frac{C}{L}}$). The amplitude and phase angles are plotted in *Figure 13-12* for several values of γ . For the underdamped cases the amplitude of $H(\omega)$ can be much larger than one near the resonant frequency of the circuit. This part of the signal frequencies is amplified, for the undamped case $\gamma = 0$ $H(\omega)$ is infinitely large. For the critically- and overdamped cases the increase in amplitude is not present anymore and the filter behaves as a low pass filter. The overshoots and ringing are not desirable in sensor systems and the value of γ should a design target, typically a value of 0.7 is a good design tradeoff between bandwidth and signal distortion.

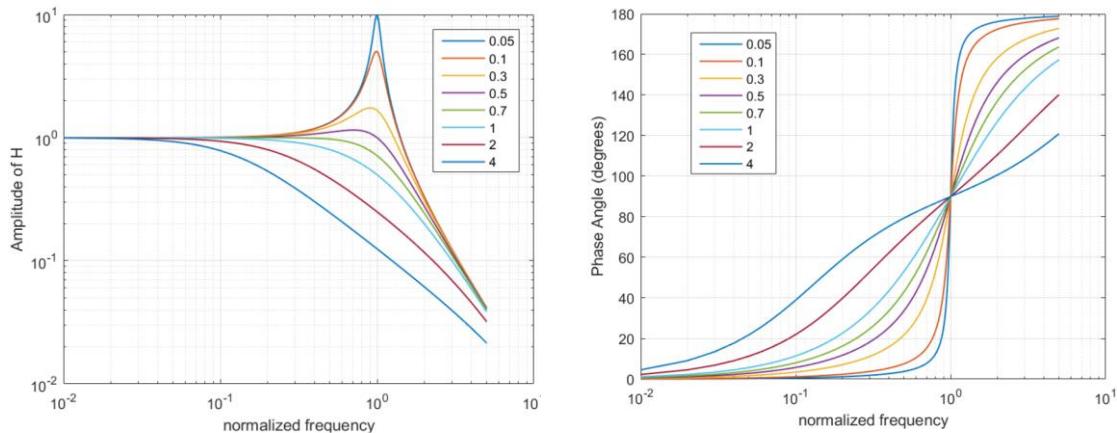


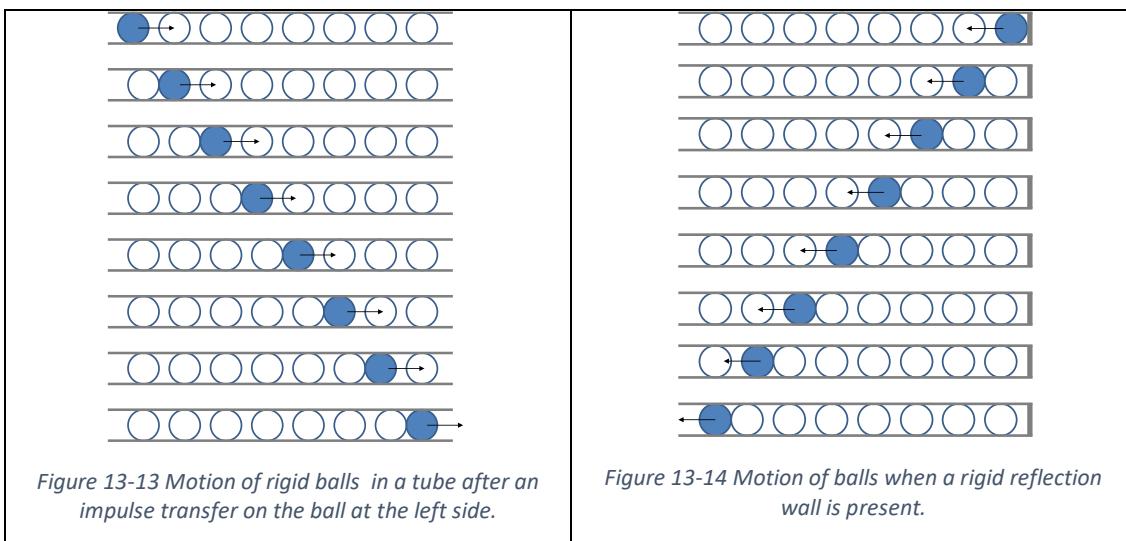
Figure 13-12 Amplitude and phase angle for a RLC filter as function of the normalized frequency (to ω_0) for various values of γ (indicated in the legend of the figure).

13.3 Transmission Line Models

For long elastic tubes, fluid transport is substantially different from that in the lumped element models. Pressure oscillations with frequency f at the input side of a tube lead to pulse waves with wavelength λ a finite pulse wave velocity (PWV). The frequency, wavelength and velocity are linked via equation $\lambda \cdot f = PWV$. For ideal tubes the PWV is independent of the oscillation frequency. Note that a constant delay time for signals with different frequency is not possible in lumped element models. Tubes can be considered short when pulse wave transit time (which is defined as tube length L over the PWV) is small compared to time scale of the signals ($1/f$). As a rule of thumb when the tube length is shorter than 10 percent of the wavelength lumped element models can be used. For longer tubes the transmission line models must be used. When pulse wave phenomena are important the fundamental equations (Maxwell in the electrical domain, Navier-Stokes in the fluid domain) or continuous distributed models have to be used. These wave phenomena in elastic tubes are very important for the physiology of the circulation and have to be taken into account. In this section the simplest electrical distributed model is used to derive the main equations. First the main features of waves are discussed.

In fluid dynamics a wave is an oscillation that transfers energy with a propagation velocity through a medium from one point to another point in space. Particles in the transmission medium are displaced from their position and transfer energy via particle interactions. The propagation velocity of energy can be much larger than that of the individual particles. (in the electrical domain pulse wave velocity is close to the velocity of light whereas the velocities of the individual electron are a fraction of this velocity). This is illustrated using [Figure 13-13](#). In this figure rigid balls are placed in regular positions with inter ball separation that is a fraction of the ball diameter. All balls are identical having the same mass and diameter. The most left-hand ball gets an impulse mv and kinetic energy $\frac{1}{2}mv^2$. When the adjacent ball is hit the ball transfers all its momentum and energy to the next ball and stops moving. This process repeats itself until the rightmost ball is hit. In this example the propagation speed of energy is much larger than that of the mass. Each ball has shifted the inter-ball separation to the right. The time it took until the last ball was hit is proportional to: 1) the number of balls minus one, i.e. seven 2) the separation between the balls and 3) inversely proportional to the velocity v . As an example, for a ball diameter of 10 cm, a ball-to ball separation of 1 cm and a velocity of 10m/s the propagation time is equal to 7 milliseconds. Hence after seven milliseconds all energy is transferred to the last ball over a distance of 87 cm. The energy wave propagation speed is approximately 125 m/s which is more than ten times larger than the maximum speed of the ball itself. The net speed of a ball is even much lower, it is only 1.4 m/s, a factor of hundred lower than that of the energy wave. After the initial push it takes 7 milliseconds before the rightmost ball can exit the tube. Note that for shorter timescales nothing happens at the outlet. An observer at the outlet would not know that a ball at the inlet side was hit.

In [Figure 13-14](#) a situation is shown where there is a rigid stop at the right side of the tube. After the most right-hand ball is hit the ball moves to the rigid wall and is reflected without energy loss. The process described above repeats itself except that the velocity of the balls and the energy wave has reversed its direction. This wave that moves towards the left side is called the reflected wave. It takes 14 milliseconds before the most left-handed ball is hit by the reflected wave. In other words, it takes 14 milliseconds before an observer at the left side of the tube knows that balls have been reflected at the right side.



This simple example shows important features of waves. The energy wave propagation speed is high, often much higher than the speed of the individual particles in the tube. Furthermore, in a long tube it takes time before the wave reaches the exit of the tube. Before that time nothing happens at the outlet side. Finally, waves can be reflected at the outlet side and the direction of motion is reversed. The phase of the forward and reflected wave are related, waves will interfere and the wave pattern in the tube can become very complex.

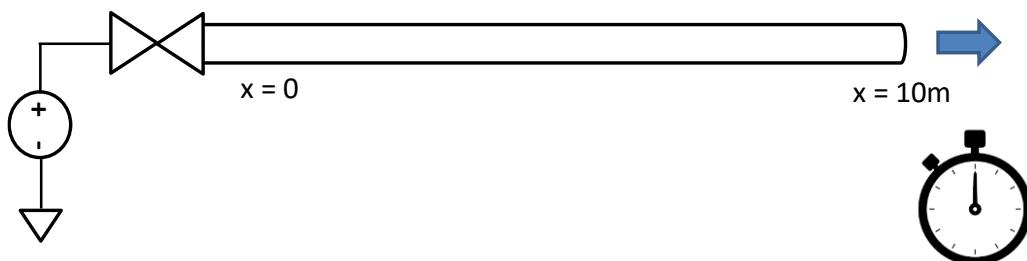


Figure 13-15 Ten-meter-long pipe with valve and pressure source at the inlet side. Time is measured between opening of the valve and first fluid leaving the tube.

Now consider fluid flow in tubes of long length i.e. 10 meters (see [Figure 13-15](#)). A rigid steel tube and an elastic rubber tube are compared, both with same diameter and length. Initially the pressure at the inlet, outlet and exterior space are identical. A valve is opened in a very short time at the inlet side, the fluid pressure at the inlet is increased. The time interval between opening of the valve and fluid flowing at the outlet side is measured, it is equal to the tube length divided by the pulse wave velocity. For the rigid pipe after approximately 20 milliseconds the pressure at the outlet changes and fluid start to flow out of the tube. For the elastic pipe it takes a few seconds before the pressure at the outlet changes and fluid starts to flow out. Hence for the rubber tube it takes one hundred times longer for the wave to reach the outlet location, the only difference being the elasticity of the tube wall. The wave velocity is approximately 10 m/s. For the rigid pipe the pressure pulse propagation velocity is of the order of 200 m/s close to the velocity of sound in air. The mechanism for wave propagation in a rigid pipe is similar to the example discussed before for the rigid balls, the balls being replaced by liquid particles i.e. molecules. For the elastic tube both fluid inertance and tube elastance play an important role. The larger the compliance of the tube to longer it takes before the pressure wave reaches the outlet. Furthermore, the inertance L is important, for instance the larger the fluid density the longer it takes before the pressure pulse reaches the outlet.

This behavior can be modeled by solving the linearized Navier-Stokes equations for fluid flow in elastic tubes but the mathematics is very complex and not much insight is gained. In this section the simplest electric analogue is used to derive the main equations for pressure and flow waves in long elastic tubes. This analogue is a transmission line (TL). In the electronic world a transmission line is used to transmit high frequency (MHz to GHz) electric signals with low losses over large distances. A well-known example is the coaxial line used in the home to transmit signals from the receiving station in the home to the television.

For the TL model (TLM) the following assumptions are made:

1. A cylindrical tube of semi-infinite length (outlet side is infinitely far away from the inlet)
2. The impedance of the tube can be modeled by an infinite chain of infinitesimal small lumped elements.
3. One dimension only (axial x-direction)
4. The tube radius r is constant
5. There are no losses in the tube and fluid (inviscid flow, tube wall material is purely elastic with a constant elastic modulus E)
6. The pressure waves are pure harmonic functions
7. Inertance L and compliance C are expressed per unit length and are constant
8. No body forces acting on the fluid

Consider a very small element of a cylindrical elastic tube at filled with fluid with flow $Q(x, t)$ and pressure $P(x, t)$ at the left side and volume V (see *Figure 13-16*). Due to changes in pressure and flow along the axial direction x the volume V may change with ΔV . The following relations relating pressure, flow and volume can be derived (equations 4-136).

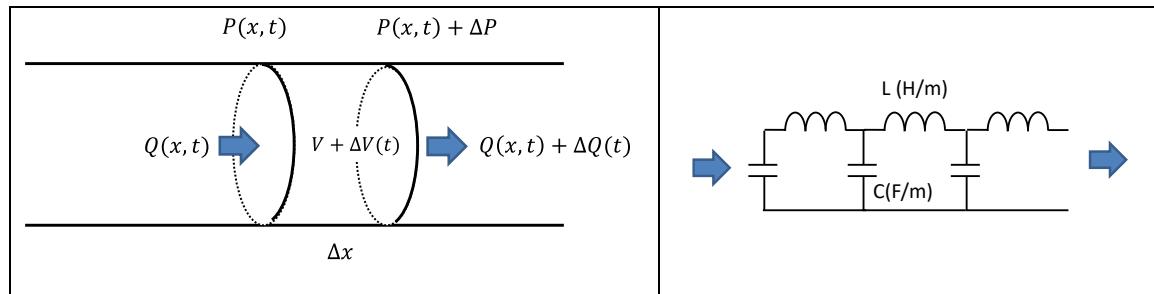


Figure 13-16: Infinitesimal element of an elastic tube of length Δx , pressure P , volumetric flow Q and volume V . C and L are specified per unit length.

The change in volume ΔV of an element with length Δx in an increment of time Δt is equal to the difference of inflow $+Q(x, t)$ and outflow $(-(Q(x, t) + \Delta Q))$ multiplied with the time increment Δt . The change in volume leads to a change in pressure ΔP due to the differential compliance per unit length times the length of the element i.e. $C\Delta x$ of the tube (equation 13.3-1). Rearranging terms gives the second difference equation (13.3-2). Finally taking the limit of infinitesimal Δt and Δx gives the differential equation 13.3-3 which links changes of flow in the axial direction with changes in pressure in time. Note that the value of the compliance and inertance in equation (4-136) are values per unit axial length. The next step is to derive an equation that links gradients in pressure in the axial direction with time.

	$\Delta V(t) = (Q(x) - (Q(x) + \Delta Q)) \cdot \Delta t = C\Delta x\Delta P$	13.3-1
	$\frac{\Delta Q}{\Delta x} = -C \frac{\Delta P}{\Delta t}$	13.3-2
	$\frac{dQ}{dx} = -C \frac{dP}{dt}$	13.3-3

Since we are considering a lossless system the only reaction force on the fluid is the inertial force. There is a pressure difference ΔP over the small tube element of length Δx , then the change in flow with time can be calculated. This gives the difference equation 13.3-4, rearranging terms gives the second equation (13.3-5) and taking the limit of infinitesimal Δt and Δx gives the differential equation 13.3-6 which links changes of pressure in the axial direction with changes in flow in time.

	$\Delta P = -L\Delta x \frac{\Delta Q}{\Delta t}$	13.3-4
	$\frac{\Delta P}{\Delta x} = -L \frac{\Delta Q}{\Delta t}$	13.3-5
	$\frac{dP}{dx} = -L \frac{dQ}{dt}$	13.3-6

Equations 13.3-3 and 13.3-6 were first derived for understanding and optimizing of telegraph lines in the 19th century and are therefore called telegraphers equations. These are coupled linear differential equations that couple flow and pressure in a tube as function of distance and time. The equations may be combined in two wave equations 13.3-7, 13.3-8 (i.e. take second order derivative of position of 13.3-3 insert 13.3-6 in the right term and rearrange) which are given below. v_p is the phase velocity (eqn 13.3-9). The parameter v_p is the phase velocity, it is the rate at which the phase of a wave propagates in space (i.e. the velocity of a specific phase point on a wave in space for instance the maximum amplitude (unit is m/s)). It appears from equation 13.3-9 that the phase velocity of the waves does not depend on the frequency of the wave, nor does it depend on the length of the tube. This implies that there is no distortion of pulse waves which can be formed by Fourier series of harmonic waves when such pulses travel through the tube. From a physical point of view the inertance of the tube limits the rate of change in flow velocity, fluid flow into an adjacent part of the tube increases the local tube volume which in turn increases the local pressure which in turn accelerates the fluid towards the forward direction and this is a continuous process. Hence kinetic energy is converted in elastic energy and so on. There is no energy dissipation. The result is a pulse wave that propagates with a constant velocity in the forward direction. The pulse wave propagates with a much larger velocity than the average flow velocity of the fluid particles.

	$\frac{d^2Q}{dt^2} = \frac{1}{LC} \frac{d^2Q}{dx^2} = v_p^2 \frac{d^2Q}{dx^2}$	13.3-7
	$\frac{d^2P}{dt^2} = \frac{1}{LC} \frac{d^2P}{dx^2} = v_p^2 \frac{d^2P}{dx^2}$	13.3-8
	$v_p = \frac{1}{\sqrt{LC}}$	13.3-9

For a pure harmonic wave (a wave with a single angular frequency ω) the steady-state pressure variations (i.e. real part of $(P(x).e^{j\omega t})$ of the equations 13.3-3, 13.3-6, 13.3-7, 13.3-8 are equations 13.3-10 to 13.3-14, j is the imaginary unit number ($j^2 = -1$).

	$\frac{dQ}{dx} = -j\omega CP(x, t)$	13.3-10
	$\frac{dP}{dx} = -j\omega LQ(x, t)$	13.3-11
	$\frac{d^2Q}{dx^2} = -\omega^2 LCQ(x, t) = \beta^2 Q(x, t)$	13.3-12

	$\frac{d^2P}{dx^2} = -\omega^2 LCP(x, t) = \beta^2 P(x, t)$	13.3-13
	$\beta = j\omega\sqrt{LC} = jk, k = \frac{\omega}{v_p} = \frac{2\pi}{\lambda} \rightarrow v_p = \frac{\omega}{k} = \frac{1}{\sqrt{LC}}$	13.3-14

The parameter β is a pure imaginary number, it is equal to jk where k is the wavenumber ($2\pi/\lambda$). Using equation 13.3-9 it can be shown that k is the ratio of the angular frequency and the phase velocity v_p (see equation 13.3-14). For waves in a loss-less ideal tube the product of the wavelength and frequency is equal to the phase velocity (i.e. $\lambda \cdot f = v_p$) and from this relationship it follows that k is equal to $\frac{2\pi}{\lambda}$ radians per unit length.

	$P(x, t) = (P_r e^{-\beta x} + P_l e^{+\beta x}) e^{j\omega t}$	13.3-15
	$Q(x, t) = (\frac{P_r}{Z_0} e^{-\beta x} - \frac{P_l}{Z_0} e^{+\beta x}) e^{j\omega t}$	13.3-16
	$Z_0 = \sqrt{\frac{L}{C}}$	13.3-17

In the loss-less case the solutions for the pressure $P(x, t)$ of the equation 13.3-13 consists of a combination of two pressure waves, the forward and reflected waves (equation 13.3-15). Since β is a pure imaginary number the wave amplitudes are constant, only the phase of the wave changes with position x . The forward wave propagates in the positive x -direction (forward with amplitude P_r) and the reflected wave propagates in the negative x -direction (reflected wave with amplitude P_l) both have the same phase velocity v_p . The amplitudes of the two waves depend on boundary conditions for pressure and flow at the inlet and outlet. In the ideal tube there is no energy loss and no damping of the waves. In the case of a semi-infinite tube there are no reflections at the outlet side and the amplitude P_l is zero.

In tubes with a finite length there are wave reflections at the outlet side. Inserting equation 13.3-15 in equation 13.3-11 gives equation 13.3-16. Note that the reflected flow wave has the minus sign because it is moving in the negative x direction. The parameter Z_0 is the characteristic impedance of the tube, it is a real number for the loss-less tube despite that the impedances of the compliance and inertance are imaginary numbers (eqn 13.3-17). Note that Z_0 does not depend on the length of the tube. The ratio between the pressure and flow amplitudes of both the forward and reflected waves are equal to the characteristic impedance Z_0 of the tube. When there are no reflections the ratio of pressure and flow at every position x is equal to Z_0 . When there are reflections this is not the case anymore. This follows from equations 13.3-15 and 13.3-16.

As mentioned before the wave behavior of pressure and flow leads to a frequency independent phase velocity of the waves, this does not occur in a lumped element model. It is therefore not allowed to model a transmission line with a simple lumped element model. Note that in the semi-infinite loss-less tube the ratio between pressure and flow is Z_0 at any location in the tube and any time, i.e. flow and pressure are in phase for all x and time t .

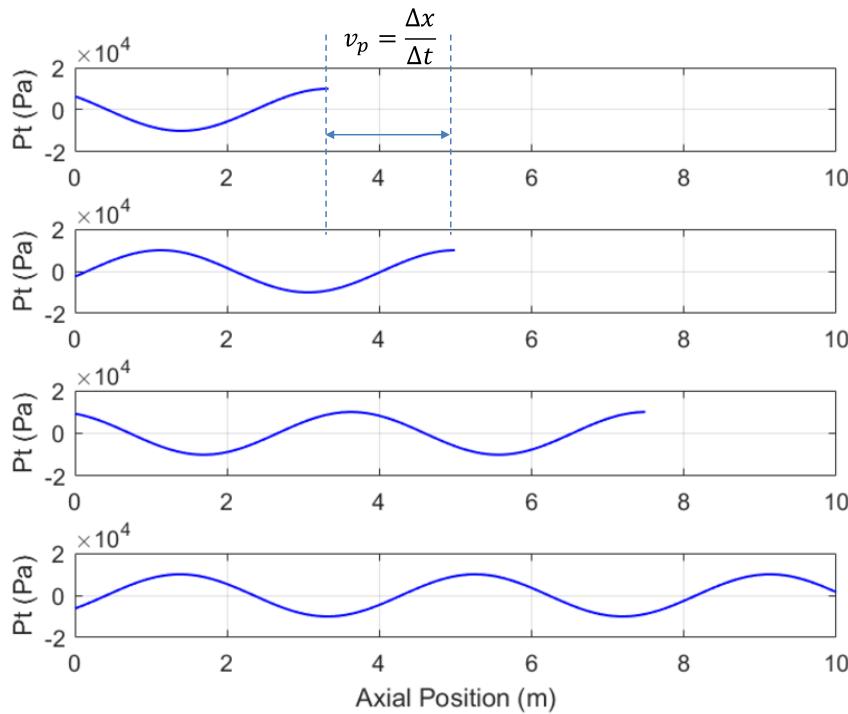


Figure 13-17 Pressure wave in a rubber tube at four times a harmonic pressure signal (1Hz, $A=10^4$ Pa) was applied at the inlet of the tube. Tube parameters: diameter 1 cm, Young Modulus: 2×10^5 N/m², fluid density: 1000 kg/m³, ratio of wall thickness to radius: 0.1, Amplitude of the forward wave: 10^4 Pa. The phase velocity is indicated in the figure, it is 3.87 m/s. The characteristic impedance Z_0 is 4.93×10^7 Pa.s/m³.

As an example of the wave propagation in a tube a harmonic pressure source with a frequency of 1Hz is connected to the tube inlet and the resulting pressure wave in the tube is calculated with equation 13.3-15 and is shown at four specific times (see Figure 13-17). The pressure is generated at the inlet (varies at $x=0$ between the minimum and maximum pressure of 10 kPa with 1 Hz) and the wave propagates with the phase velocity v_p in the positive x-direction. The phase velocity (i.e. the velocity of a specific phase point on the wave) is 3.87 m/s. The peak flow is equal to P_r/Z_0 and is equal to 203 ml/s. The flow and pressure waves are in phase.

The foregoing model a semi-infinite tube was assumed, real tubes have finite length. This leads wave reflections at the tube ends (see Figure 13-18), i.e. P_l is not zero more. As a result, complicated pressure and flow waves are generated and the tube impedance depends on position x. The first parameter that needs to be known is the wave reflection coefficient σ which is the fraction of the right moving wave that is reflected in the opposite direction at the outlet of the tube ($x=L$).

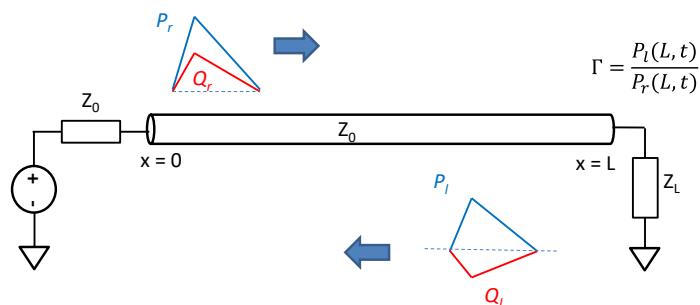


Figure 13-18: Schematic drawing of a tube with an inlet pressure source that is matched to the characteristic impedance and a load impedance Z_L at the outlet. Forward and reverse waves and the pressure reflection coefficient Γ are indicated in the figure.

The reflection coefficient is defined in equation 13.3-18. The pressure reflection coefficient Γ and the propagation constant $\beta = jk$ are introduced, as discussed before for the loss-less tube β is a pure imaginary number. The impedance at the outlet ($x=L$) is equal to the load impedance Z_L and the ratio of pressure and flow is determined by Z_L . Therefore at $x=L$ equation 13.3-15 can be written as equation 13.3-19. Rearranging terms and using 13.3-18 gives the pressure reflection coefficient (equation 13.3-20). Inserting equations 13.3-18 into 13.3-15 gives equation 13.3-21 this shows that the tube impedance depends on position x when there are reflected waves.

	$\sigma = \frac{P_l e^{j k L}}{P_r e^{-j k L}} = \frac{P_l}{P_r} e^{2 j k L} = \Gamma e^{2 \beta L}$	13.3-18
	$Z_L = \frac{P_r e^{-\beta L} + P_l e^{\beta L}}{\frac{P_r}{Z_0} e^{-\beta L} - \frac{P_r}{Z_0} e^{\beta L}}$	13.3-19
	$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0}$	13.3-20
	$Z(x) = Z_0 \frac{1 + \Gamma e^{-2\beta(L-x)}}{1 - \Gamma e^{-2\beta(L-x)}}$	13.3-21

The pressure reflection coefficient depends on the load impedance and tube characteristic impedance. The forward and reflected running waves have fixed phase relation and amplitude ratio and interfere. In some cases (specific tube length and frequency) the interference leads to the formation of standing waves. A few interesting load conditions are:

1. $Z_L = Z_0$. The load impedance is equal to the characteristic impedance, the matched condition. The reflection constant is zero and the impedance of the tube is equal to the characteristic impedance at all positions x , at the inlet side the tube behaves exactly as the semi-infinite length tube, there are no reflections. This is the conditions that is used most often for the electrical networks.
2. $Z_L = 0$. The load is zero, hence $P_L = 0$, the shorted condition. The pressure reflection coefficient is equal to minus one ($\Gamma = -1$). The reflected pressure wave amplitude has the negative amplitude of the forward wave, the reflected current wave has a positive sign.
3. $Z_0 = \infty$. The load is infinitely large, hence $Q_L = 0$, the open condition. The reflection constant is equal to one ($\Gamma = 1$). The reflected pressure amplitude is equal to the forward amplitude.

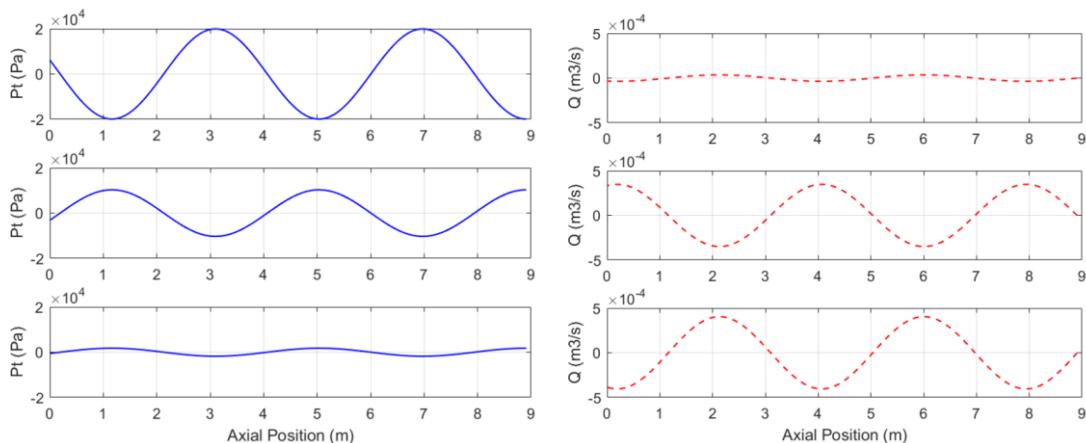


Figure 13-19 Steady state total pressure (left) and flow waves (right) for three random times for the tube of Figure 13-17 with an open outlet after reflections. The tube length was 9 meters.

The total pressure $P(x,t)$ and flow waves $Q(x,t)$ for a tube at three arbitrary moments in time are shown in [Figure 13-19](#). The reflection coefficient Γ is one. The tube length is chosen such that standing waves do not occur. The flow at the outlet is always equal to zero ($Z_L = \infty$). The pressure and flow waves have no regular shape, depend on time and the tube impedance $Z(x)$ can vary between zero and infinity depending on position and time. When reflections occur the situation is complex, pressures at the inlet side may differ from the applied pressures and flow cannot be simply derived by measuring the pressure in the tube. Information on the reflected wave is required before understanding of the flow and pressure waves is complete.

Of particular relevance for the circulatory physiology is the input impedance of the tube as function of the tube length L (i.e. $(Z_{in} = Z(x = 0))$ and load Z_L . There are a few interesting cases (we use the parameter γ , for the loss less tube it is equal to β).

1. Length L is equal to a half wavelength (i.e. $\gamma L = n\pi$). The input impedance is equal to the load impedance ($Z_{in} = Z_L$).
2. The length of the tube is equal to a quarter wavelength and multiples (i.e. $\gamma L = \frac{\pi}{2} + n\pi$). The input impedance is equal to: $Z_{in} = Z_0^2/Z_L$.
3. Matched load. The input impedance is equal to Z_0 .
4. The load is a short ($Z_L = 0, P_L = 0$). Then the input impedance is purely imaginary and is equal to: $Z_{in} = jZ_0 \tan(\gamma L)$. The input impedance can vary be between zero and infinity ($\gamma L = \frac{\pi}{2} + n\pi$).
5. The load is an open ($Z_L = \infty, Q_L = 0$). Then the input impedance is purely imaginary and is equal to: $Z_{in} = -jZ_0 \cot(\gamma L)$. For $\gamma L = n\pi$ the input impedance is infinite. For $\gamma L = \frac{\pi}{2} + n\pi$ the input impedance is zero.

For all other cases the input impedance at the source side depends is a complicated manner on the tube length L , characteristic impedance Z_0 and the load impedance Z_L . The tube impedance is not constant anymore, the impedance depends on position x .

13.3.1 Transmission line with losses

The models discussed above are for an ideal case, there were no losses in the tube. Realistic tubes have losses, the models for a tube with losses are discussed briefly below. The telegraphers' equations for lossy tubes are derived using the infinitesimal model shown in [Figure 13-20](#). The infinitesimal element of the tube contains four elements.

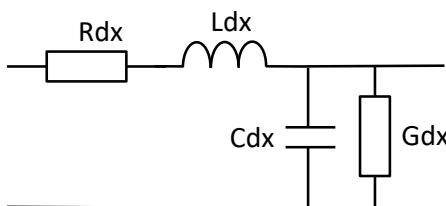


Figure 13-20 Four element model of an infinitesimal segment of a tube with losses. Note that R, L, C, G are numbers per unit length.

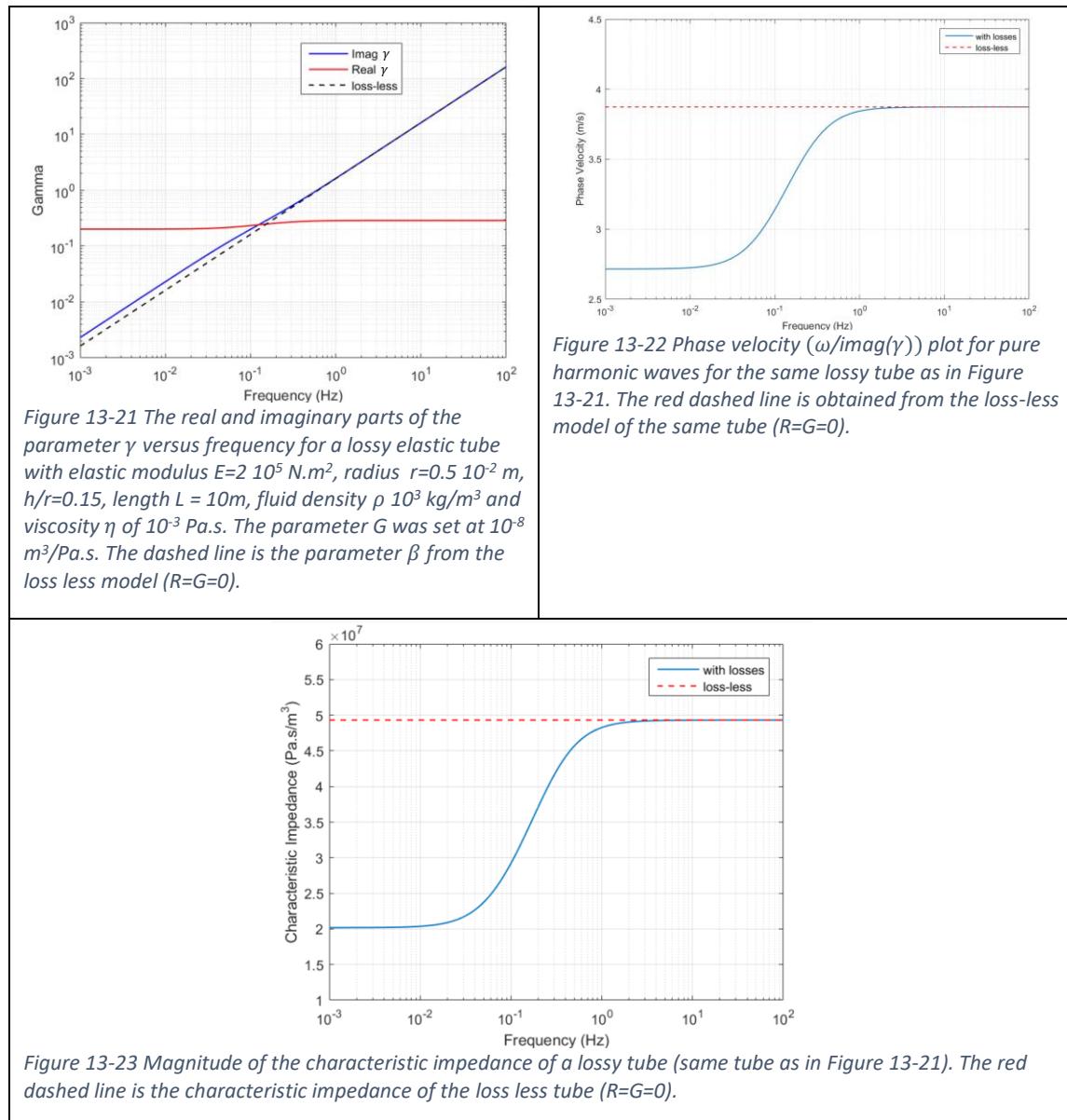
The resistance R is the flow resistance due to the viscosity of the fluid. The conductance G is related to losses in the visco-elastic material of the tube wall. Both R and G can be frequency dependent. The telegrapher's equation and wave equations for arbitrary wave forms are given by equations 13.3-22 to 13.3-25. The telegrapher's equations (13.3-24, 13.3-25) and wave equations (13.3-28, 13.3-29) for harmonic waves now include terms with R and G .

	$\frac{dP}{dx} = -(L \frac{dQ}{dt} + R.Q)$	13.3-22
	$\frac{dQ}{dx} = -(C \frac{dP}{dt} + G.P)$	13.3-23
	$\frac{d^2P}{dx^2} = LC \frac{d^2P}{dt^2} + (LG + RC) \frac{dP}{dt} + GR.P$	13.3-24
	$\frac{d^2Q}{dx^2} = LC \frac{d^2Q}{dt^2} + (LG + RC) \frac{dQ}{dt} + GR.Q$	13.3-25
	$\frac{dP(x)}{dx} = -(R + j\omega L).Q(x)$	13.3-26
	$\frac{dQ(x)}{dx} = -(G + j\omega C).P(x)$	13.3-27
	$\frac{d^2P(x)}{dx^2} = \gamma^2 P(x)$	13.3-28
	$\frac{d^2Q(x)}{dx^2} = \gamma^2 Q(x)$	13.3-29
	$P(x, t) = (P_r e^{-\gamma x} + P_l e^{+\gamma x}) e^{j\omega t}$	13.3-30
	$Q(x, t) = (\frac{P_r}{Z_0} e^{-\gamma x} - \frac{P_l}{Z_0} e^{+\gamma x}) e^{j\omega t}$	13.3-31
	$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} = \alpha + j\beta$	13.3-32
	$v_p = \frac{\omega}{\beta}, \beta = \text{Im}\{\gamma\}, v_p = f(\omega)$	13.3-33
	$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$	13.3-34

The parameter γ (equation 13.3-32) replaces the parameter β (equation 13.3-14), it has both a real and imaginary part. The pressure waves are now damped waves, the amplitude decreases when the wave travels through the tube.

The imaginary and real part of γ are plotted in Figure 13-21 for a rubber tube with material parameters characteristic of water and rubber and a diameter of 1 cm. The imaginary part of γ depends on frequency in a different way than the parameter β of the loss-less tube, hence the phase velocity $v_g = \omega/\text{Imag}(\gamma)$ depends on frequency (equation 4-165). As a result harmonic waves with different frequency travel with different phase velocity through the tube. This will lead to distortion of the shape of pressure pulses which consist of a Fourier series of harmonic waves. The dependence of the phase velocity of pure harmonic waves on frequency is shown in Figure 13-22. For low frequency there is a large deviation from the loss-less frequency independent value, for higher frequencies the phase velocities of lossy and loss-less tube values are almost the same. The damping parameter α is approximately 0.15 m^{-1} . This would correspond with loss factor of $e^{-0.15}$. This would correspond to a damping loss of 14% per meter length. The longer the tube the larger the distortion and the larger the damping. The characteristic impedance Z_0 can be obtained by differentiating equation 13.3-30 to position x and inserting the result in equation 13.3-26. This gives an equation for the characteristic impedance Z_0 that has both a real and imaginary part (equation 13.3-34). The frequency dependence of the magnitude of Z_0 is plotted in Figure 13-23. At low frequencies the amplitude is a factor of two smaller than at high frequencies. The characteristic impedance of the tube merges with the value of the loss-less tube for higher frequencies. At high frequencies the tube parameters γ , Z_0 and v_p are similar to the loss-less tube as the imaginary terms become more

important due to the increase of the angular frequency. The damping parameter α is sufficiently small for this tube but can become more prominent for tubes with smaller diameter. The example is for a tube where losses are not very large, however for smaller diameter tubes the resistive losses become more prominent, the resistance will dominate the inertance and the tube will behave more as a low pass RC filter. The high frequency components are filtered out.



13.3.2 Summary TLM appendix

1. When the tube length is larger than 10 percent of the wavelength of the pulse a distributed transmission line must be used to model pressure and flow waves in the tube.
2. Both forward and reflected waves are solutions of the wave equations.
3. A loss-less tube has a characteristic impedance $Z_0 = \sqrt{\frac{L}{C}}$ which is a real number and it does not depend on tube length L. L and C are the inertance and compliance per unit length.
4. For a loss less, tube the phase velocity $v_p = \frac{1}{\sqrt{LC}}$ and does not depend on frequency.
5. The pulse transit time is equal to $PTT = \frac{l}{v_p}$, l is the length of the tube.
6. The ratio of the amplitudes of the reflected and forward wave of a tube with load Z_L at the outlet of the tube is $\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0}$.
7. When the tube is terminated by a load that is equal to the characteristic impedance there are no reflections and the tube impedance is equal to the characteristic impedance at all positions x.
8. For other loads there are reflections and the tube impedance depend on the position x.
9. For a tube with losses the phase velocity of harmonic waves depends on frequency and waves are damped. Pressure pulses that are formed by a Fourier series are distorted.

14 Appendix Chapter 5

14.1 The electromagnetic flowmeter

A conductor moving through a magnetic field generates an electric potential. This principle, also employed in e.g. microphones and generators, is the basis for the electromagnetic blood velocity meter. This device does not measure the flow but the velocity of the blood. If we also know the cross-section of the tube through which the blood flows, and if we also known the flow profile (or assume that it is parabolic), we can compute the blood flow.

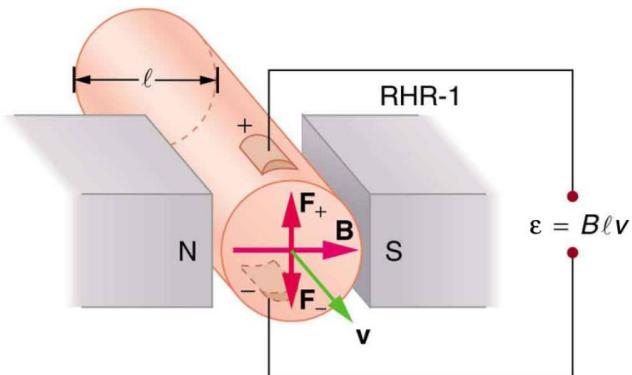


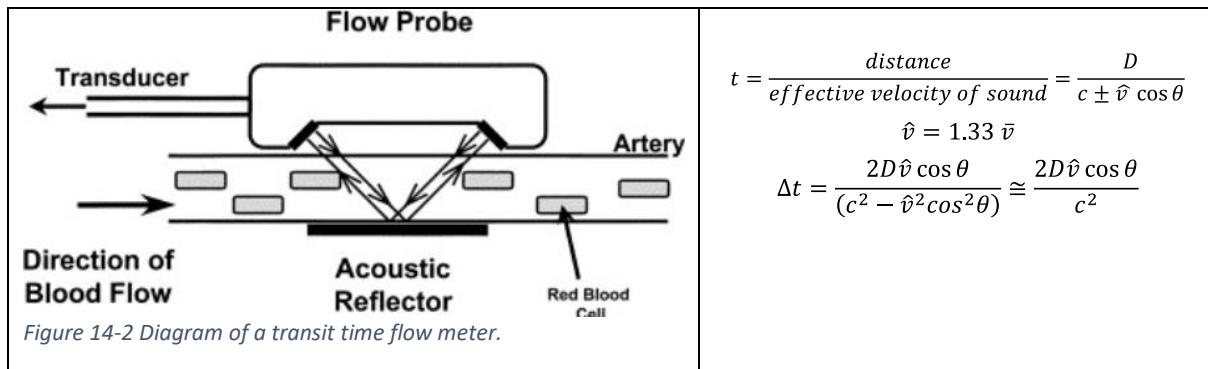
Figure 14-1 In the electromagnetic blood velocity meter, a conductor (blood) moving through a magnetic field B generates an electric potential E proportional to the average blood velocity.

In this application, the conductor is blood, the magnetic field B is externally applied, and the electric potential E is measured on the inside surface of the non-conducting, non-magnetic tube (Figure 14-1). E can be shown to be the average of the velocities of the blood in the tube. The application can be expanded to cases where the vessel wall is conductive. The advantage is that now potential E can be measured on the outside of the vessel wall.

This technique is useful when the blood flow can be measured outside the body, e.g. in a heart-lung machine, which temporarily takes over the function of the heart. It is not practical for measuring the flow in intact blood vessels, where the major problem is that access to the vessel (and thus surgery) is required to position the inductor that generates B and the electrodes that measure E . Miniature devices have, however, been designed which can be put on the tip of a catheter.

14.2 Time Transit Flow Meter

A diagram of a transit time flow meter is shown in Figure 14-2. It consists of a housing with two transducers mounted at distance D and angle θ and an US reflector at the other side of the artery. The transmitter and receiver function of a transducer is being changed with a repetition frequency of a few hundred Hertz. The transit time depends on the direction (upstream or downstream) of the ultrasound wave. The transit time and difference in transit time is indicated next to Figure 14-2. From this time difference the average flow (\bar{v}) can be obtained. The difference in transit time is in the order of nanoseconds. This transit time measurement requires complex high frequency electronics. The accuracy is very good, timing errors are the main source of error. The other parameters are fixed and known by design.



15 Appendix Chapter 9 – Therapeutic Devices

First the impact of positive pressure ventilation on hemodynamic parameters is discussed.

15.1 Impact of Positive Pressure Mechanical Ventilation on Hemodynamic Parameters

The heart, large blood vessels and lungs have a strong mechanical interaction. The lung pressure influences the transmural pressures of the heart chambers and the large blood vessels in the thorax via the intrapleural space pressure. This interaction is shown in Figure 15-1. In Figure 15-2 a plot of measured airway pressure and beat to beat stroke volume is shown during a positive pressure flow-volume controlled ventilation. A modulation of the stroke volume is observed, the stroke volume is maximum at the end the inspiration period, the minimum occurs during the expiration period. When the lung pressure is increased during inspiration the pleural pressure and transpulmonary pressure increase and this triggers a cascade of steps (see Figure 15-3).

The effect of increased pleural pressure on cardiac output can be described best using the Guyton model of circulation. In short the increased pleural pressure leads to changes in the right and left ventricular stroke volumes due to the following effects. (LV: left ventricle, RV: right ventricle)

- Increase in right atrium pressure
- Reduction of venous return due to a decrease in pressure difference with the periphery
- Right ventricular preload reduction
- Increased pulmonary resistance due to reduction in pulmonary blood vessel diameter
- Increased right ventricular afterload
- Reduced RV stroke volume
- Delayed by pulmonary transit time to LV, reduced LV stroke volume after 3-5 beats during the expiration period.
- During inspiration blood is squeezed from the left pulmonary vein into the left ventricle.
- The LV preload increases during the inspiration period
- The LV afterload decreases due to an decrease in wall stress and smaller pressure of the extra-thoracic arteries.
- The LV stroke volume increases during PPV inspiration period

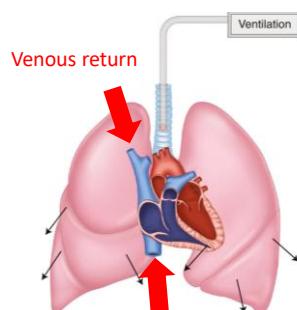


Figure 15-1 Schematic drawing of the heart and lungs in the thorax and venous return reduction.

Hence the increased pleural, lung and transpulmonary pressures lead to a reduction and modulation of the stroke volumes. At larger inspiratory and PEEP pressures this reduction is very large. This reduces pulmonary blood flow and subsequently oxygen delivery to the tissues. Thus the high pressures required during PPV ventilation of patients with respiratory failure (ARDS, COVID, acute lung injury) causes a reduction of blood flow and blood oxygenation. This reduction must be balanced against the gain in ventilation. In the end an optimum ventilation-perfusion ratio must be obtained. This balance requires the highest level of invasive monitoring and skill and experience of ICU clinicians. The impact on lung ventilation, lung injury and hemodynamics was summarized, it is now possible to discuss the advantages and disadvantages of pressure controlled and flow-volume controlled ventilation.

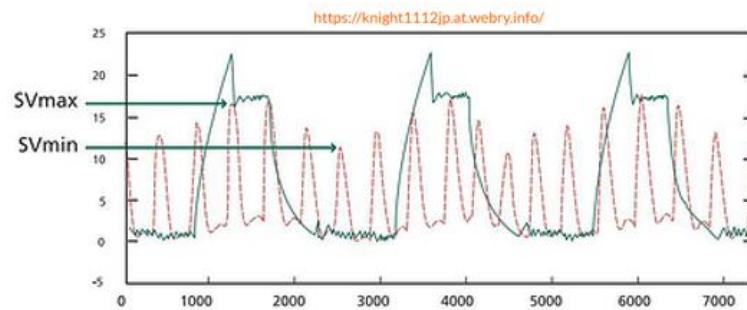


Figure 15-2 Measured airway pressure (green line) and stroke volume (red dashed line) during volume control PPV. (from: source indicated on top of the figure)

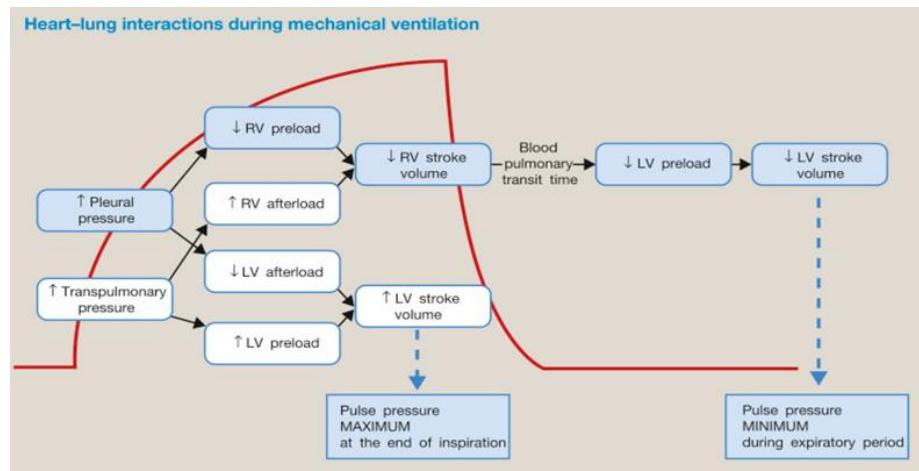


Figure 15-3 Cascade of effects on right and left ventricular preload and afterload during PPV. Source F. Michard et al. Crit. Care; 4;2000; pp 282-289

A high level description of the (intelligent) ASV and NAVA closed loop mechanical ventilation systems are described in the following sections.

15.2 Adaptive Support Ventilation (ASV)

In a conventional pressure controlled mode of ventilation clinicians set ventilator parameters like tidal volume, pressures, breaths per minute and inspiratory and expiratory ventilation time to achieve clinical therapeutic goals such as alveolar ventilation and oxygenation (see Figure 15-4). These settings have to be regularly updated to optimize ventilation for the patient. The ASV closed-loop mechanical ventilation method automates the update of ventilation settings for both controlled mandatory ventilation for passive patients and pressure support mode for active patients. Furthermore lung protective ventilation is a main target in the ASV algorithms. The ASV algorithm was developed on the basis of the minimum work of breathing models from Otis and Mead (see Figure 15-7, [1],[2]) to limit mechanical power delivered to the patient.

Three ventilator settings are needed to start ASV and control ventilation and oxygenation (see Figure 15-4). These are the mode, i.e. ASV mode, the minimum minute volume in percent of the population average tidal volume (for CO₂ elimination) and the PEEP level and fraction of oxygen in the inspired gas (FiO₂) in percent (for oxygenation).



Figure 15-4 Ventilation control settings in conventional MV and in the ASV mode. (source Hamilton)

For spontaneous breathing patients ASV is the breathing rate is controlled by the patient. ASV is now a volume targeted pressure support mode with automatic adjustment of the level of pressure support according to the respiratory rate. The pressure support level is adjusted to maintain minute volume. When the patient recovers and the breathing strength increases the pressure support level is automatically reduced, the minute volume is held constant. This is useful for weaning of the patient from the mechanical ventilator. Triggering and cycling off of the ASV pressure support is the same as in the PSV mode.

ASV requires four settings (see Figure 15-5). The first is the height and gender of the patient. The 100% minute volume is calculated by a 100 ml gas mixture per kilogram of the predicted body weight (PBW). PBW depends on the length and gender of the patient, it is not the measured weight of the patient (i.e. correct for obesity). A higher or lower percentage of this MV can be chosen depending on the patient characteristics. From the demographic parameters the anatomic dead space volume can also be estimated. This is needed to determine the lowest tidal volume given the minute volume and breathing rate limitations. In step 3 the parameters related to oxygenation are set, these are PEEP and percentage of oxygen in inspired air. Finally the controls for synchronization (trigger & cycling levels) and lung protective ventilation (inspiratory pressure P ASV limit) are set. In controlled ventilation the tidal volume is given by the ratio of minute volume and breathing rate, this is the minute volume curve. The minute volume curve, the patient target zone (red box) and the target point are shown in Figure 15-6. This plot is shown on the monitor display of the ventilator.

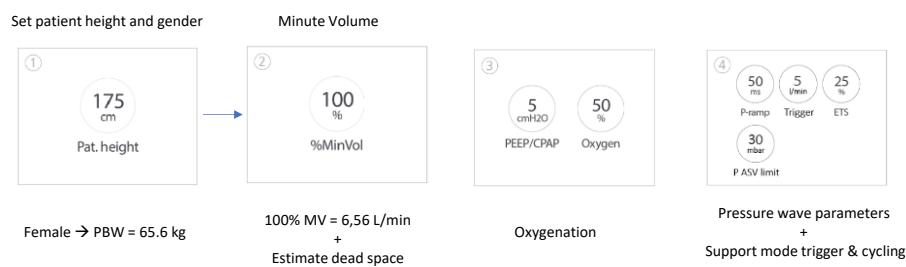


Figure 15-5 Input settings for the ASV mode of ventilation (source Hamilton)

For passive patients ASV is a volume-targeted pressure-controlled mode of ventilation with automatic adjustment of the inspiratory pressure, respiratory rate and inspiratory expiratory time ratio (I:E ratio). The minute volume is set by the user and a tidal volume (TV) versus respiration rate (RR) curve can be constructed (see Figure 15-6). The expiration is passive and in most cases the

expiratory flow has an exponential shape and decrease as function of time. The expiratory time constant (τ_{exp}) is measured breath by breath (time needed for a decrease of the flow of 1/e or 37% of maximum flow). A typical value is 0.6-0.8 seconds for a normal lung. The expiratory time should be at least three times larger than the expiratory time constant to avoid hyperinflation and autoPEEP. Since the respiration rate and expiratory time are known the inspiratory time can be calculated. The target tidal volume is obtained by adjusting the inspiratory pressure. The maximum tidal volume is controlled by setting the maximum inspiratory pressure.

For a passive patient parameters of the lung and airway can be estimated from the measured pressure, flow and tidal volume waves. The inspiratory and expiratory resistance to flow and the static compliance can be estimated from the “equation of motion 15.2-1”. P_{tot} is the pressure measured at the inspiratory port of the ventilator, Q is the airflow measured at the proximal location and V_t is the tidal volume obtained by integrating flow versus time. R is the airway resistance including the parasitic resistances of the breathing set tubes and endotracheal tube and C_{st} is the static lung chest wall compliance. When it is assumed that R and C_{st} are constants the actual values of these parameters can be determined from the measured flow and pressure waves using a least square technique [3]. The extracted values for the inspiratory and expiratory phases of the breath are available in the ASV user interface screen and give important information to user on the state of the lung and airways. Typical values for R and C_{st} are 10 cmH₂O.s/L and 60 ml/cmH₂O. When the values of the lung mechanical parameters are available the breathing rate that corresponds with the minimum work per breath (WOB) can be calculated using the Otis and Mead equations [1],[2]. This is the actual target RR rate (see Figure 15-7). It corresponds with the estimated resting spontaneous breathing rate of the patient. It is assumed that this rate minimizes the mechanical power dissipated in the patient. For a normal lung this breathing rate is approximately 15 breaths per minute. This is the target breathing rate. It can be modified when the expiratory time constant changes.

$$P_{tot}(t) = PEEP + R \cdot Q(t) + \frac{V_t(t)}{C_{st}}$$

15.2-1

ASV continuously adjusts ventilator settings to remain within the target zone, i.e. the safety frame (see Figure 15-6). A lung protective strategy limits maximum tidal volumes (maximum tidal volume of the box). Note that the maximum inspiratory pressure limits the maximum tidal volume. Too low alveolar ventilation is avoided by a minimum value of the tidal volume, the lower limit of TV of the box. The high RR limit of the box is set to avoid dynamic hyperinflation of the lung, it follows from the minimum value for the expiratory time and minimum value of the inspiratory time (determined by Pmax). The low RR limit of the safety box follows from Apnea rules and limit of tidal volumes. The shape of the box gives information on the patient archetype and status of the lung and airways.

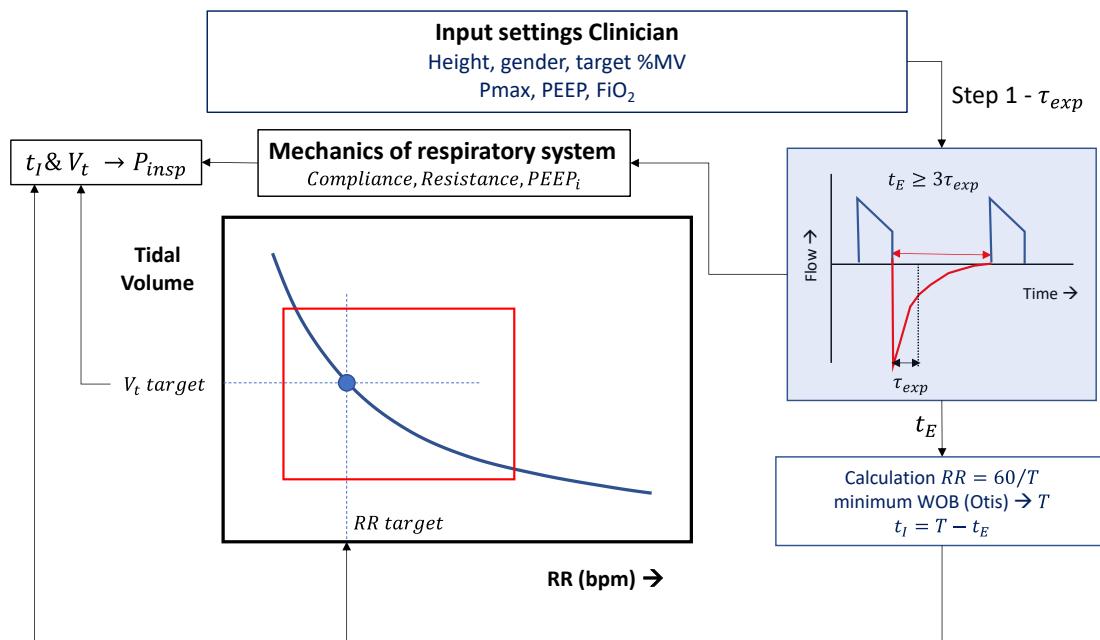


Figure 15-6 Tidal volume versus breathing rate curve determined by set value of %MV with safety frame (red square), the target point for tidal volume and respiration rate is the blue dot.

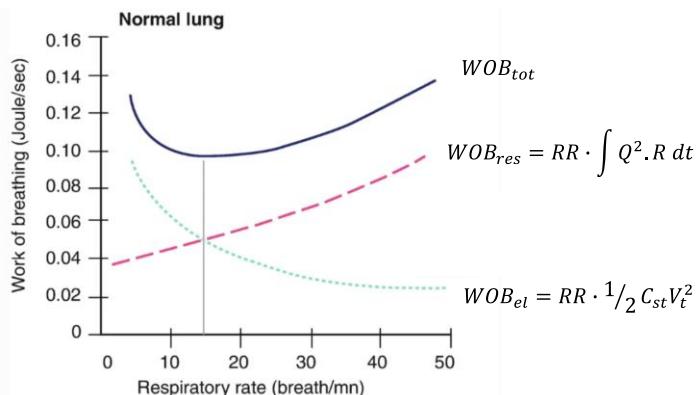


Figure 15-7 Work of breathing versus the breathing frequency for a normal lung . The total WOB is the sum of elastic and resistive components (from [Arnal, Schaedler and Kiralki,[4]])

ASV automatically detects spontaneous breaths and switches over to a volume targeted mode of pressure support ventilation. The respiration rate is now controlled by the patient. The tidal volume is controlled by variation of the inspiratory pressure. When the patient recovers the pressure support level is reduced. After a while the patient condition and monitor parameters indicate that a weaning test can be started. ASV has an automatic mode to optimize the weaning process. ASV was introduced in 1998 as a commercial product. Minute ventilation and oxygenation were set by the user and were not updated.

The effectiveness, safety and efficiency of ASV has been the topic of several smaller clinical studies. ASV has been compared to the conventional therapy, it was at least comparable in outcome, no superiority in treatment has been reported. The efficiency, i.e. time spent by clinicians to optimize treatment was reduced.

A new version of ASV with automated updates of oxygenation and CO₂ elimination was introduced in the products around 2010. It is the Intelligent ASV mode, it is briefly discussed in the next session.

15.3 Intelligent ASV

The next step is to automate ventilation beyond ASV and include automation using measured oxygenation (pulse oximetry SpO₂) and CO₂ elimination (end-tidal CO₂ pressure). The user has to set targets and ranges for pETCO₂ and SpO₂. INTELLIVENT-ASV automatically and continuously adjusts respiratory rate, tidal volume, inspiratory time, PEEP, and Oxygen depending on physiologic input from the patient (PetCO₂, SpO₂, lung mechanics, spontaneous breathing). The clinician only needs to adjust target ranges for PetCO₂ and SpO₂ based on the patient condition, and weaning strategy. The ASV algorithm is used for the control of the ventilation volumes, pressures and rate.

A diagram of the control loops for a mandatory ventilation using INTELLIVENT-ASV is shown in Figure 15-8.

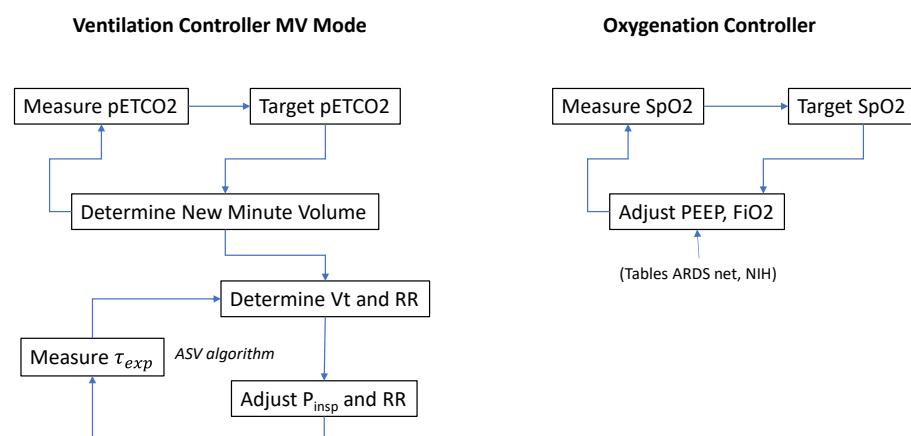


Figure 15-8 Control loops for mandatory ventilation using INTELLIVENT-ASV.

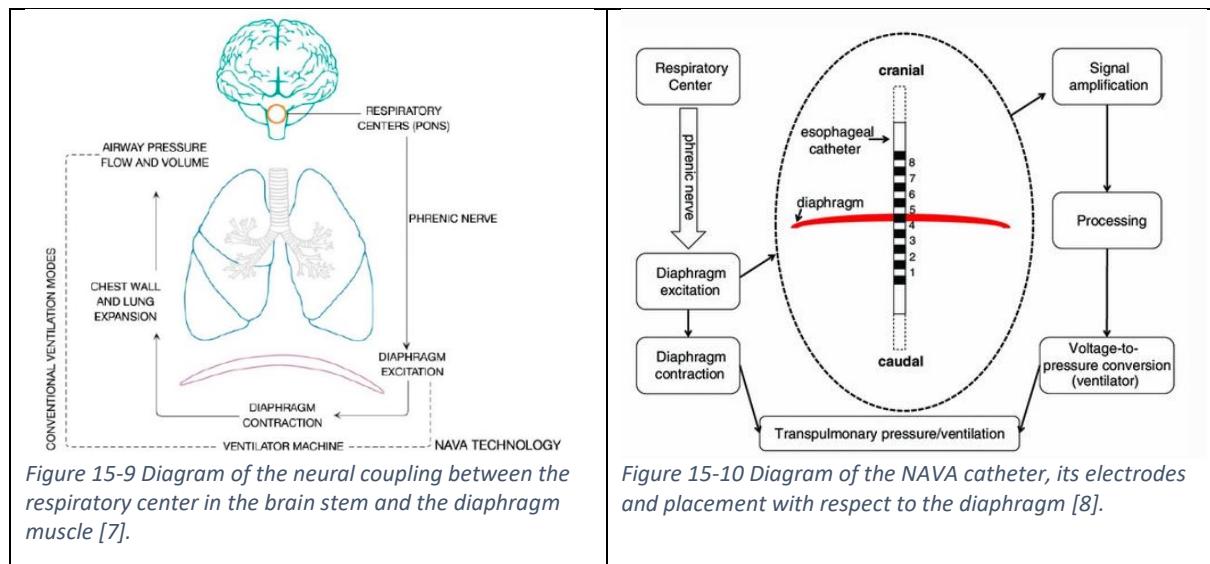
The minute volume control uses the measured end-tidal CO₂, it is compared to the target set value and the new minute volume is determined (i.e. in proportion to the difference of the two values). For instance when the measured ETCO₂ is larger than the set value the minute volume is increased. This new minute volume serves as input for the ASV controller and the tidal volume, respiration rate and inspiratory pressure are adjusted. The expiratory time constant is measured breath to breath and ventilation parameters are automatically adjusted when needed.

The oxygenation controller compares the measured and target SpO₂ values and the variables that are changed are PEEP and fraction of oxygen (FiO₂) in the inspired air. The values for PEEP and SpO₂ are copied from recommendations of the ARDS net and the NIH organization [5]. For instance when SpO₂ is too low for the given PEEP and FiO₂ the next recommended level of these parameters is chosen. Smaller studies show that this method is as safe and effective as conventional therapy [6]. As for ASV there are no large clinical studies available that give data on outcome, i.e. superiority of the technology over conventional therapy in terms of mortality, complications length of stay in the ICU. Future large studies are needed to test efficacy.

15.4 Neurally Adjusted Ventilatory Assist (NAVA)

During mechanical ventilation spontaneous breathing must be detected and supported by the ventilator. Spontaneous breathing can be both beneficial but can also have drawbacks for patients. It is difficult to synchronize the ventilator and estimate the support level given a certain patient effort. Note that the level of support must be carefully chosen. Asynchronies and too low or too high level of support have been correlated with length of stay in the ICU and even with increased mortality. NAVA is an assist mode where the electrical activity of the diaphragm is measured and is used to detect the effort, synchronize the ventilator with the effort and to determine the level of support.

NAVA requires the use of a special catheter with 8 electrodes, it can be combined with a nasal feeding tube that is often needed for ICU patients (see Figure 15-9, Figure 15-10, Figure 15-11). Correct placement is very important and can be a challenge.



The respiratory center in the brain stem controls the respiration effort. The effort is mainly determined by the partial CO₂ pressure and pH of arterial blood in the brain. During respiratory distress the pH is low and respiratory drive is large. The respiratory center generates a signal in the phrenic nerve that triggers the respiratory muscles by generating an electric EMG signal, it is in the order of 10's of μ V. The diaphragm is the main respiratory muscle and is used in NAVA. As noted before the electrical muscle activity is proportional to the respiratory drive and eventually the mechanical strength of the contraction. Therefore NAVA measures the electrical myographic signal of this muscle.

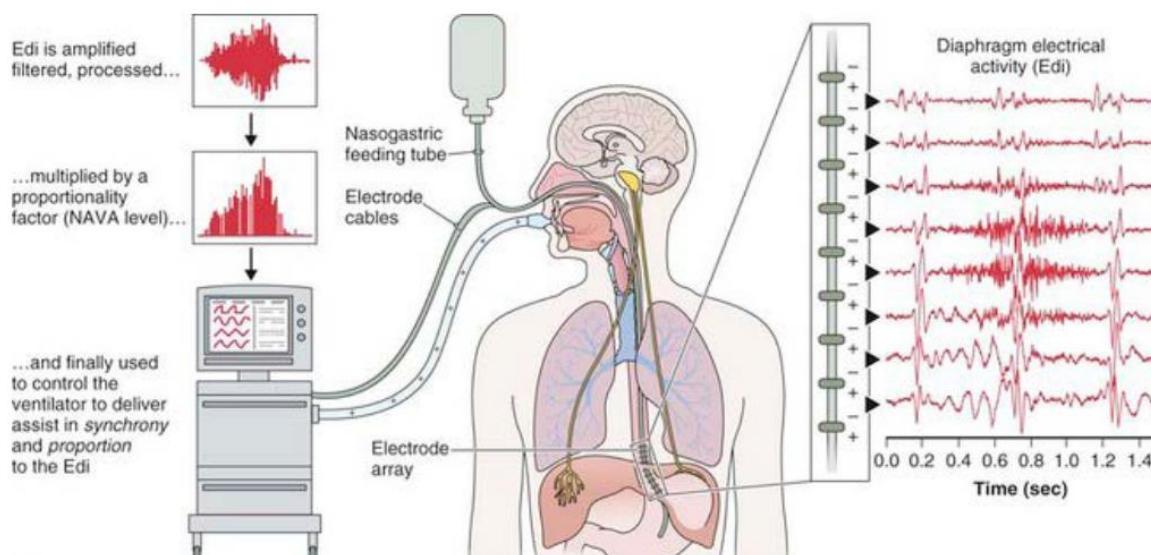


Figure 15-11 Diagram of the NAVA technology integrated in a ventilator and electrode signals (<http://clinicalgate.com/invasive-mechanical-ventilation/>).

The heart is close to the diaphragm and the ECG signals disturb the smaller NAVA signals. In Figure 15-11 a diagram of the NAVA system and the signals from the eight electrodes are shown. The electrode signals can be used to optimize placement of the catheter by looking to the ECG and NAVA

signals and by using anatomic information. Note that the frequencies of the EMG signal are larger than that of the ECG signal. The ECG signal is also acquired. Using the signals from the eight electrodes the electrode signals are processed by filtering and digital signal processing. The processed EMG signal (EAdi) is obtained. This signal is integrated and is called EAdi (Electrical Activity of the diaphragm). The EAdi signal is multiplied by a proportionality factor to convert it to muscle pressure in cmH₂O. The EAdi signal is used to trigger and cycle the inspiratory support of the ventilator. NAVA is a so-called proportionally assist ventilation method (PAV). The inspiratory pressure delivered by the ventilator during inspiration is given by the following relation:

$$P_{aw} = NAVAlevel * EAdi + PEEP \quad 15.4-1$$

Thus the ventilator inspiratory pressure is directly proportional to the NAVA EAdi signal, a strong effort is accompanied by a strong pressure support. Typical pressure, flow, volume and EAdi waveforms obtained during NAVA are shown in Figure 15-12 [7]. Note the proportionality of the airway pressure to the EAdi signal. Triggering is done when the EAdi signal passes a threshold level ($\sim 1 \mu V$). Cycling off is done during the falling edge of the EAdi signal at a level of 70% of the maximum amplitude.

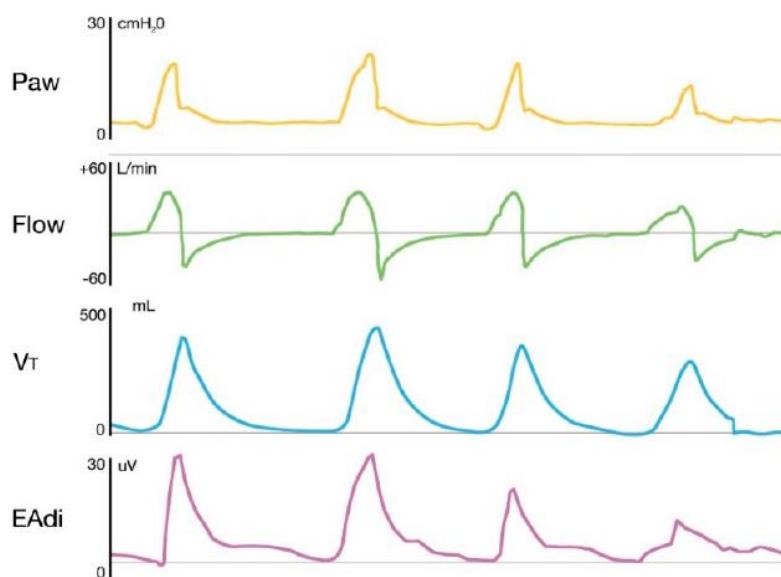


Figure 15-12 pressure, flow, volume and EAdi signal during NAVA support ventilation [7].

The claimed advantages of NAVA assist ventilation are synchronization with respiratory effort (reduced number of asynchronies), triggering is not affected by autoPEEP, effects of leakage on triggering are reduced, it follows irregular breathing patterns, assist is in proportion to the respiratory drive of the patient, reduced risk of overdistention and protection of the diaphragm muscle. Furthermore the average airway pressure is lower this could reduce VILI. However lung protective ventilation may be an issue. A strong effort is accompanied by a strong pressure support, this could result in too large tidal volumes which could cause VILI.

15.5 Infusion Systems [8]

Syringes are used to rapidly provide the patient with a known volume of one of a large variety of drugs. For fast action, drugs can be injected into the bloodstream, usually intravenously in the large veins of the arm, hand or subclavian vein (see Figure 15-13). For slower drug action, intramuscular injections can be given; their disadvantage is that the rate at which the drug becomes effective will depend on where the bolus was deposited in the tissues and on usually unknown characteristics of the patient, such as muscle bulk.

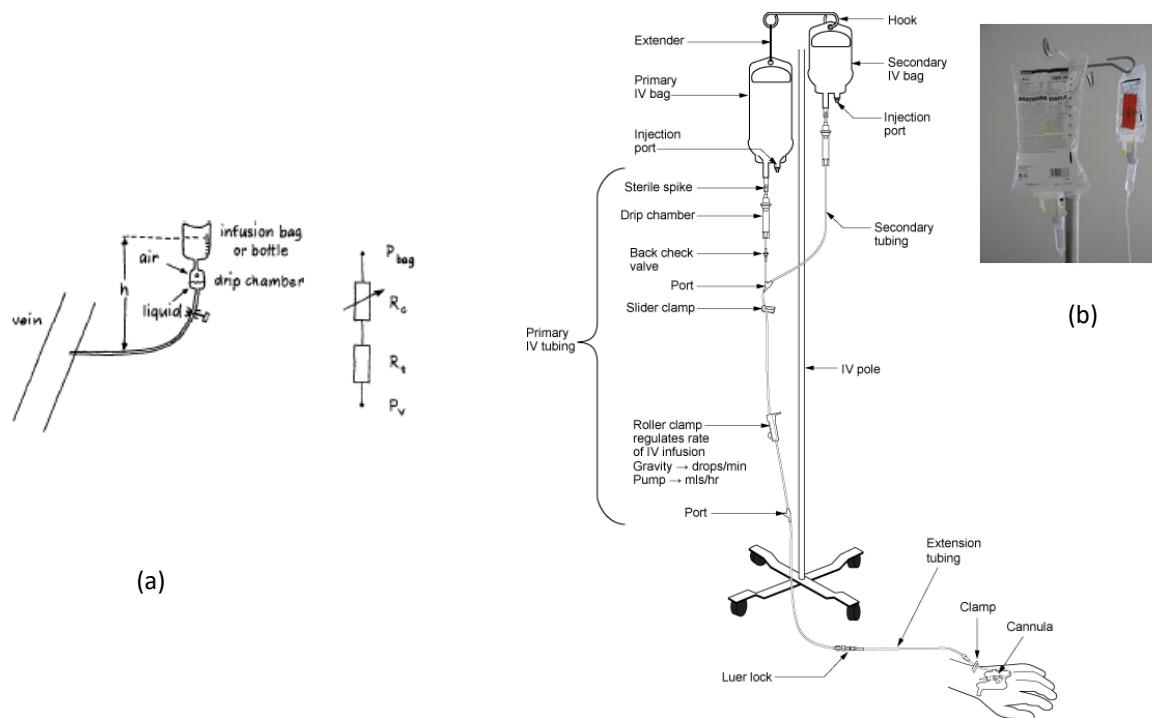


Figure 15-13 Infusion drip system and its model (a). The drip chamber contains a volume of air, so that one can see individual drops of infusion fluid fall. If an infusion is delivered by a drip system, the infusion flow is often given as number of drops per minute. Diagram of a complete IV infusion system (b).

Although an injection can directly influence a measurement, their indirect influence-through the action of the drug-is much more important. Since many drugs influence the values of the measured variables, a correct interpretation of the measurements is possible only if it is known when and which drugs have been applied and in what dosage. Providing this information to a monitoring system is important for reliable record keeping, but especially when "intelligent" alarms are generated. In that case, the information must be provided to the system *at the time the injection is given*. Although bar codes on syringes and bar code readers are sometimes used, the requirement that no extra effort is to be demanded from the clinician makes this approach unpopular and error-prone.

Infusion drips realize a slow, yet carefully controlled drug flow. A bag or bottle, usually filled with a physiological salt solution to which a known volume of the drug is added, empties itself slowly through a catheter into the patient, often into a vein. The desired flow rate is adjusted through a restriction in the catheter and can be read off as a number of drops in a drip chamber (Figure 15-13 (a)). The following analysis will show that this method is not reliable.

In order to compute the flow rate, the model of Figure 15-13(a) can be used, which assumes infusion into a vein. The pressure in the bag P_{bag} is equal to the hydrostatic pressure ρgh , where ρ is the fluid's density (which may be assumed equal to that of water), h is the height of the infusion fluid level above the site where the fluid enters the vein, and g is the gravitation constant. The pressure in the vessel is P_v . R_c stands for the resistance of the infusion line, which can be varied by adjusting the flow restriction. R_t stands for the resistance of the tissues; in an intravenous infusion R_t can often be neglected. By varying R_c , the infusion flow Q_{inf} is adjusted to a value:

$$Q_{inf} \propto (P_{bag} - P_v)/(R_c + R_t)$$

15.5-1

We see, however, that flow will change if P_v or R_t changes. If this can be the case, the flow rate cannot be accurately maintained.

An infusion flow *into* the patients requires that P_{bag} is larger than the pressure inside the vessel P_v . A lower value of P_{bag} would cause backflow of blood into the infusion line. Infusion into a *vein* can be realized by suspending the infusion bag at a certain height above the vessel. Infusion into an *artery* cannot be done this way. Since an arterial pressure of 200 mm Hg corresponds to a water column of 266 cm high, the height of the bag would have to be almost 3 m above the blood vessel. Few rooms have that height. In arterial infusions, the infusion bag is therefore enclosed in an extra inflatable bag, which is pressurized to 200 or 300 mmHg. The infusion flow rate depends on fluctuations of the arterial pressure. In arterial infusions, a high pressure difference $P_{bag} - P_v$ may exist; an extremely reliable, constant value of the flow restriction R_c must prevent accidental infusions of too large flows.

Infusion pumps can realize accurately known flow rates. Figure 15-14 shows some pump types. In a roller pump, one of its wheels moves the point of collapse in a flexible tube, thus pushing the fluid in the tube forward. In a centrifugal pump, the spin of the rotor creates a centrifugal force that drives the fluid toward the walls and the outlet; fluid enters along the axis to replace the fluid that is pumped out. A volumetric pump has two cycles. In a rapid cycle, the pump's plunger downward movement fills the pump's chamber with infusion fluid from the bag. In a slow cycle, the plunger moves upward and the chamber's fluid is infused into the patient. Unidirectional valves allow flow in one direction only. In a syringe pump, a (large) syringe is slowly emptied into the patient.

A syringe pump is used most frequently as it is very reliable and its contents can be inspected. Infusion pumps need a high degree of safety as injection of a too large dose or the lack of injection can lead to grave consequences to the patient. Therefore modern syringe pumps are equipped with connectivity options such that the injection process can be monitored and archived.

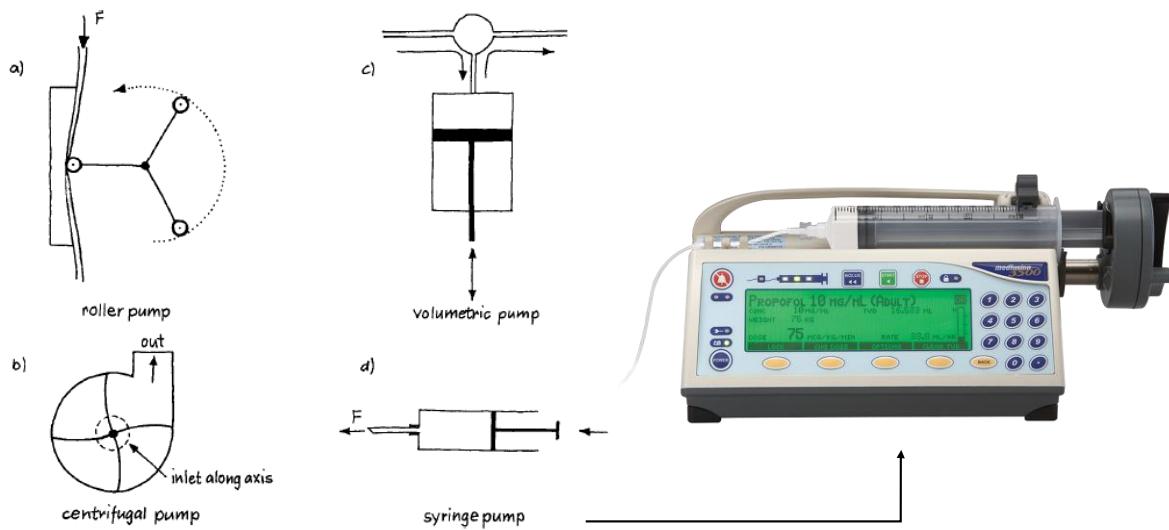


Figure 15-14 Roller pump (a), centrifugal pump (b), volumetric pump (c), syringe pump (d). A syringe pump (right) is most often used in an anesthesia system.

All these pumps can be modelled as flow sources, where the flow that they generate is independent of resistances and counter pressures. High pressures arise when the line is partially or fully occluded.

The pressure in the line is therefore often measured at the infusion pump; it can be used to generate an alarm and/or to switch off the pump if the pressure becomes too high.

15.6 Defibrillators and Cardio Pulmonary Resuscitation

Cardiac arrest is a condition when the heart has no mechanical output, no blood is pumped into the circulation and there is no spontaneous breathing. When there is no oxygenated blood flow to brain, heart and other organs irreversible damage will occur and this leads to death within a few minutes.

15.6.1 Types of cardiac arrest

There are three major types of cardiac arrest (see Figure 15-15). Ventricular tachycardia and ventricular fibrillation lead to uncoordinated multi-site electrical stimulation of the ventricular tissue and the heart muscle quivers and there is no ejection of blood. The cardiac output vanishes within a few seconds. During fibrillation the muscle is still active and a local oxygen and nutrients are depleted in a few minutes. The heart becomes a stiff muscle and is not capable of contractions anymore. Fibrillation can be converted to a normal rhythm by a short duration large voltage / large current electrical impulse (1000V/10A, 10 ms) though the heart tissue. When the event lasts for a few minutes and the heart muscle is stiff, defibrillation will have a low probability of success.

Therefore Cardio-Pulmonary-Resuscitation (CPR) is needed when the time-to-arrest is larger than a minute. During CPR forceful chest compressions and artificial ventilation cause a small net flow of oxygenated blood that will perfuse heart and brain and will increase the probability defibrillation success and may reduce neurologic brain damage. Defibrillators and CPR are discussed in more detail in following sections. Presently about 20% of the cardiac arrests are due to VT/VF, in the 1970's and 1980's this was 70%-80%.

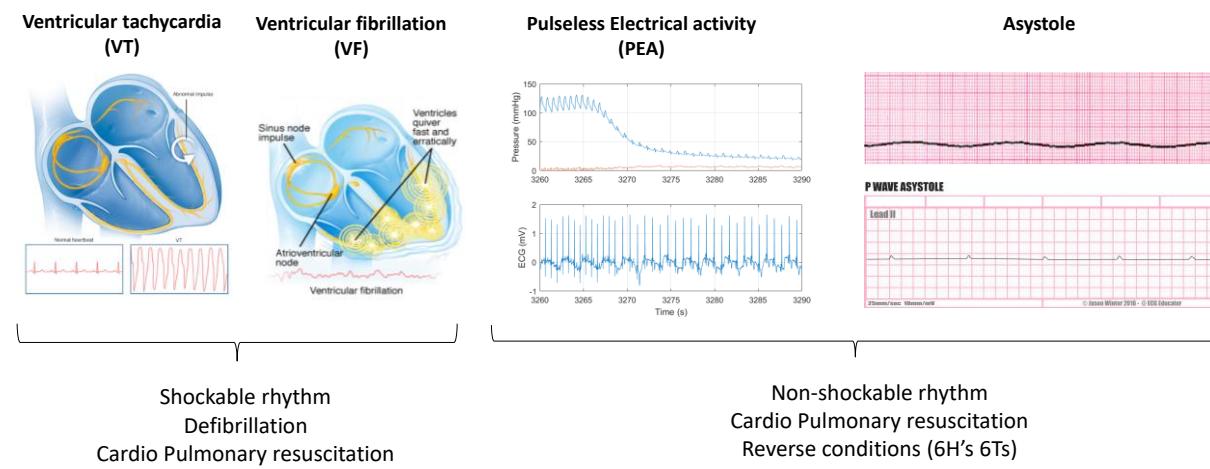


Figure 15-15. The three main types of cardiac arrest: left ventricular tachycardia and ventricular fibrillations (the shockable rhythms), middle pulse less electrical activity and right asystole.

Pulseless electrical activity (PEA) has become a common type of arrest. The heart is still electrical active and should have a pulsatile perfusing rhythm. The electrical stimulation does not lead to an effective contraction of the heart muscle and cardiac output is low or even absent. The first thing is to apply CPR to buy time for further treatment to revert the cause of the arrest. Possible causes are abbreviated with the acronym 6H-6T where H stands for hypo or hyper and T for Toxin or Tamponade. Some examples are too low or too high potassium ion concentrations, pulmonary embolisms (hypo perfusion), hypoxemia (drowning), drug overdose, cardiac tamponade. Defibrillation is of no use and even harmful. The probability of survival is lower than that of VT/VF.

The third rhythm is asystole, there is no electrical nor mechanical activity. This rhythm is often the end-phase of a cardiac arrest due to VT/VF or PEA. The arrest has lasted a considerable time and brain, heart and organ damage is severe, the probability of survival is very low, much lower than that of the other two rhythms.

15.6.2 Defibrillator

During ventricular tachycardia and ventricular fibrillation there are one or multiple locations in the ventricles that cause uncoordinated contractions of the heart. The depolarization of the heart fibers in the ventricles is desynchronized. A defibrillator sends a short large current pulse through the heart muscle and this pulse depolarizes all muscle cells in the heart. In many cases fibrillation stops and the normal rhythm comes back. This is called return of spontaneous circulation (ROSC). The probability of shock success decreases exponentially with the time passed since the onset of the event and on the condition of the heart (hypoxia, metabolism, size and location of an infarct). The probability of shock success depends also on the waveform and energy of the pulse and location of the electrodes. The electric field over the heart muscle must be minimally 3V/cm – 5V/cm but must be smaller than 60V/cm to avoid damage of the heart. The total current must be in the ampere range. Both voltage and current are important. The mechanism is not completely understood.

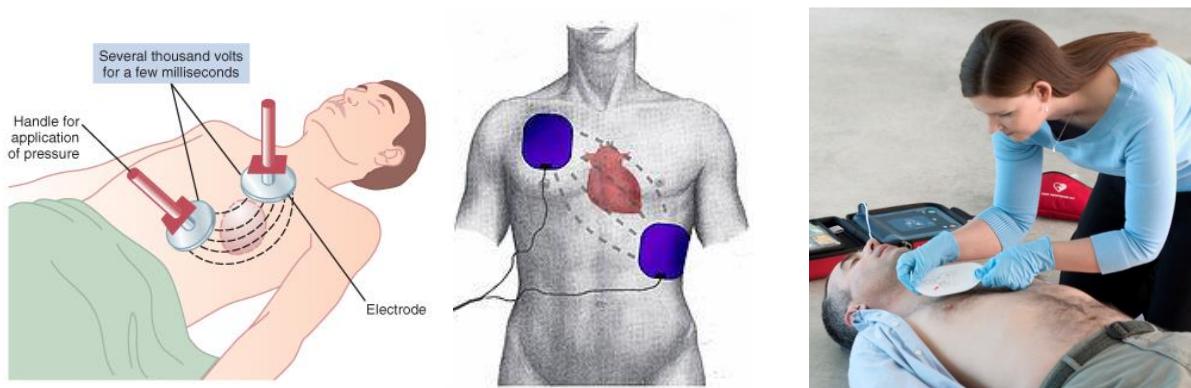


Figure 15-16 Electrodes and electrode placement before defibrillation.

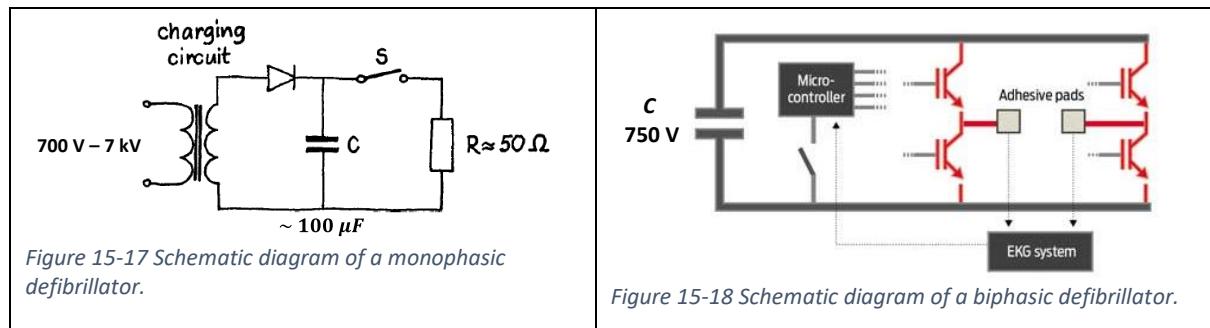
A maximum current in the order of 10 to 20 Amperes must pass through the skin. Therefore electrode size and contact resistance are very important. In Figure 15-16 the electrodes and the placement of electrodes is shown for a system with paddles and for gel electrodes. Paddles are pushed firmly on locations indicated in the figure. Paddles were the dominant electrodes in older systems but are now more and more large Ag/AgCl gel electrodes are used (right figure). These electrodes have lower contact resistance, cause less burns and are safer for the user and patient. It is important that the current flows through the heart muscle, electrodes must be placed on the left side of the ribcage and on the upper right part of the chest (middle figure).

In its simplest form a defibrillator is a large capacitor that is charged to a high voltage and is discharged by closing a switch (see Figure 15-17). The energy stored in the capacitor is equal to:

$$E = \frac{1}{2} CV^2 \quad 15.6-1$$

For 1kV and 100 μF capacitor the stored energy is equal to 50 J. When the capacitor is discharged by a simple switch the pulse is monophasic, i.e. has one polarity. It appeared that an energy of 300 J to 350 J was needed for a high probability of shock success for a monophasic pulse. This is a lot of energy and it can damage the heart muscle, especially when repeated shocks are needed. For a

monophasic device voltages in the order of 2kV to 3kV are used. This increases the size of the device.



It was found that so-called biphasic pulses had a higher shock success at much lower voltage and energy of the pulse. Bi-phasic pulses with an energy of 150J at a voltage of 750V have higher success rates and cause less harm to the heart. A bi-phasic pulse cannot be generated with the simple defibrillator shown in Figure 15-17 and electronic switches are needed. The schematics of a modern biphasic defibrillator is shown in Figure 15-18. An H-bridge high voltage electronic switch can be used to reverse the polarity of the pulse. The devices include a microcontroller and also include an ECG measurement such that the user can analyze the ECG rhythm and can detect whether the shock was successful. In Figure 15-19 two modern defibrillation devices are shown.

Automatic External Defibrillator (AED)



Monitor-Defibrillator



Figure 15-19 Philips FRx AED and Philips MRx monitor defibrillator.

An AED is an automatic external defibrillator that also guides the user during when it is switched on. It has a display to show images, sometimes ECG and has voice prompts. The capacitor is charged automatically, the ECG signal is analyzed and the user is advised when a shock can be given. There are AED's for the public space that can be used by layman. More advanced versions such as the FRx are designed for professional use in ambulances and hospitals. For professional use by experts in code teams the defibrillator settings can be applied manually and shocks can be administered at wish. The monitor function is used to guide the resuscitation effort and adjust when needed.

Such devices are also used for cardioversion, i.e. the conversion of atrial fibrillation into a normal sinus mode of atrial contractions. In this case the shock needs to be synchronized with the R-peak ECG pulse. When the shock is administered at the wrong moment in the cardiac cycle VF could be triggered.

There are also internal implantable defibrillators for patients with high risk for cardiac arrest. This function is most often combined with a pacemaker function. Lower voltages and energies can be

used as electrodes are in direct contact with the heart muscle. The device can also be used for cardioversion of AF. These devices are known as implantable cardioverter defibrillators (ICD).

It was found that shock success decreases strongly with time after the onset of the arrest. A few minutes may already reduce the chance of success with a large factor (7%-10% reduction in survival per minute). Since in many cases the time between detection and application of defibrillator is longer than a few minutes the heart highly hypoxic and stiff. The probability of arrest survival can be improved by application of cardio pulmonary resuscitation (CPR) where artificial oxygenated blood flow is generated and the heart, brain and other organs are perfused. CPR buys time for further treatment.

15.7 Cardio Pulmonary Resuscitation (CPR)

Cardio Pulmonary Resuscitation (CPR) was pioneered in the 1950's and 1960's in the United States by Kouwenhoven and Safar. The basic principle is illustrated in Figure 15-20.

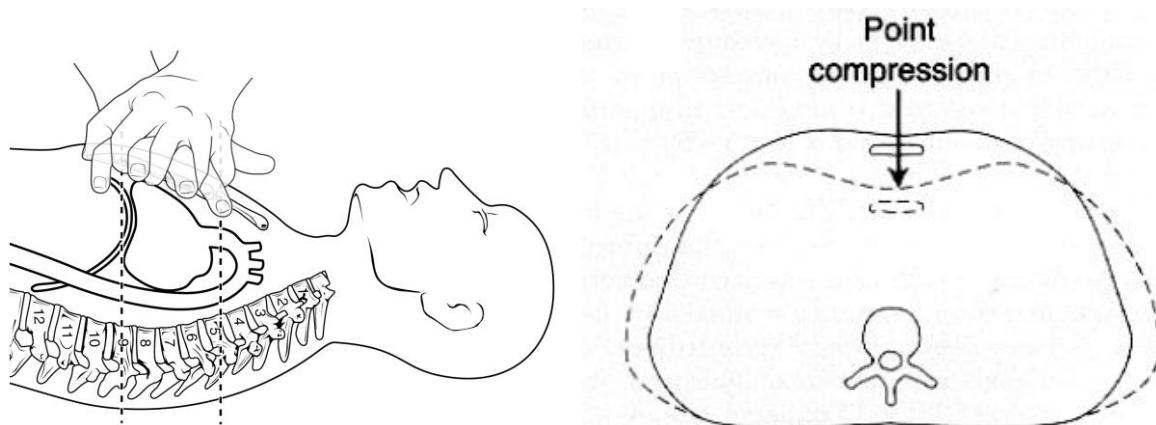


Figure 15-20 Schematic drawings of chest compressions during CPR.

Both hands are placed on the lower part (above the sternal notch) of the breast bone and chest compressions with a depth around 4 cm to 6 cm are administered at a rate between 80 and 120 compressions per minute. The compressions reduce the thoracic volume and may compress the heart between the sternum and spine. There are several mechanisms proposed how blood flow is generated during CPR. Whatever the mechanism is of blood flow generation the flow to the brain and the coronary vessels of the heart is only a fraction (~20 % - 30 %) of the flow at normal physiologic conditions. Note that this holds when CPR is done perfectly. The quality of CPR has been studied extensively and it has proven to be difficult to do prolonged high quality CPR and rescuer fatigue is another issue. The reduced blood flow generated by CPR is just sufficient to protect the brain and heart from ischemic damage for a short period (a few minutes) and to improve the condition of the heart by perfusing with oxygenated blood. The probability of survival decreases exponentially with a rate of 7%-10% when no CPR is applied and between 2% and 5% when CPR is applied. CPR buys time for treatment to revert the cause of the arrest either by defibrillation or other medical treatment.

The physiology of blood flow during CPR is still poorly understood. Two different models have been proposed (see Figure 15-21). In the cardiac pump model it is assumed that the heart is compressed between the sternum and spine and blood is forced out of the heart. During the recoil phase the pressure in the heart drops and new blood flows into the heart. This model is similar as the model for

normal physiologic blood ejection out of the heart and one expects a similar pressure and flow waveforms albeit with smaller magnitude.

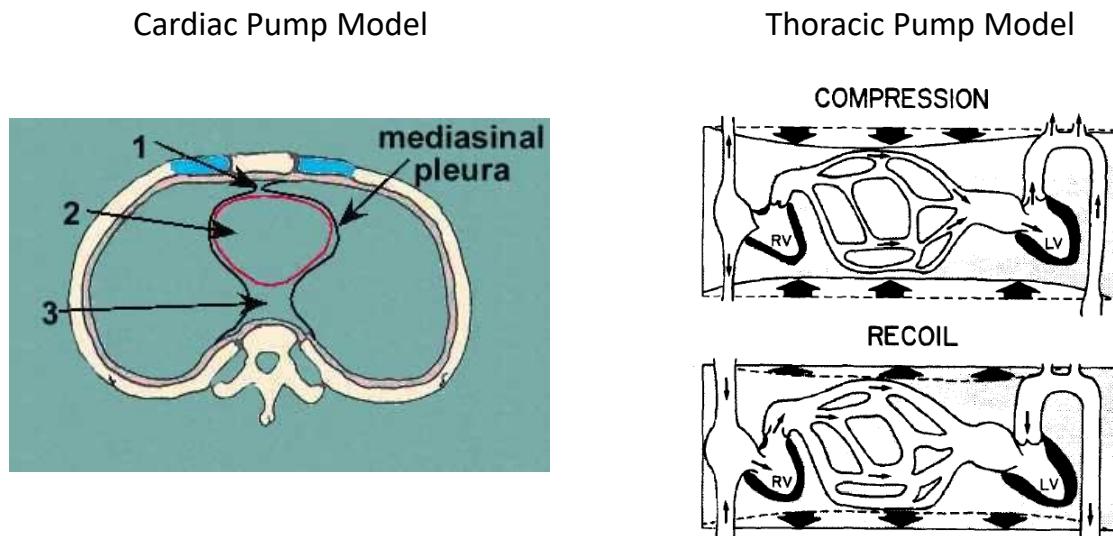


Figure 15-21 Schematic drawing of cardiac pump model and thoracic pump model.

From an anatomic viewpoint it is not clear how the heart can be compressed between the sternum and spine, the distances are too large. It was observed during catheter procedures that coughing can generate large pressures and forward flows. This observation has led to the thoracic pump model (right part Figure 15-21). During a chest compression the intrathoracic pressure will rise uniformly and external pressures are exerted on the large blood vessels, the heart and the pulmonary vessels. When the valves are functional net flow can be generated. During compression blood in the pulmonary vessels and the left heart is forced via the left heart into aorta. The valves in the right heart are closed (backwards flow is not possible). During chest recoil the aortic valve closes, the valves in the right heart open and blood flows into the pulmonary vascular system and left heart. In the large blood vessels blood flow will now have both a forward flow component (compression phase) and reverse flow component (recoil phase). This reduces the net flow with factor of 4 to 5. Both models are overly simplistic but offer a first order interpretation of the physiology of blood flow during CPR.

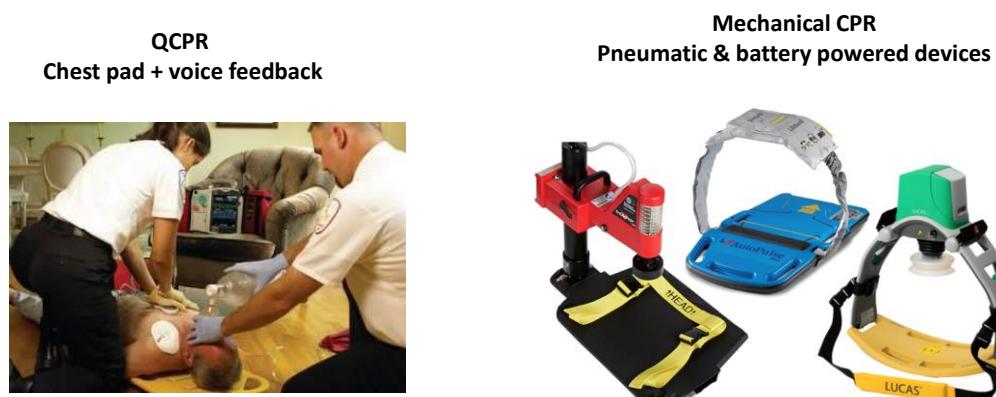


Figure 15-22 Options to improve manual CPR (QCPR) and mechanical CPR devices.

CPR is difficult to perform well and there are options to improve quality of manual CPR. Both assist devices and mechanical powered devices are available to improve or guarantee CPR quality. Some options are shown in Figure 15-22. To improve manual CPR the chest compression depth and rate are

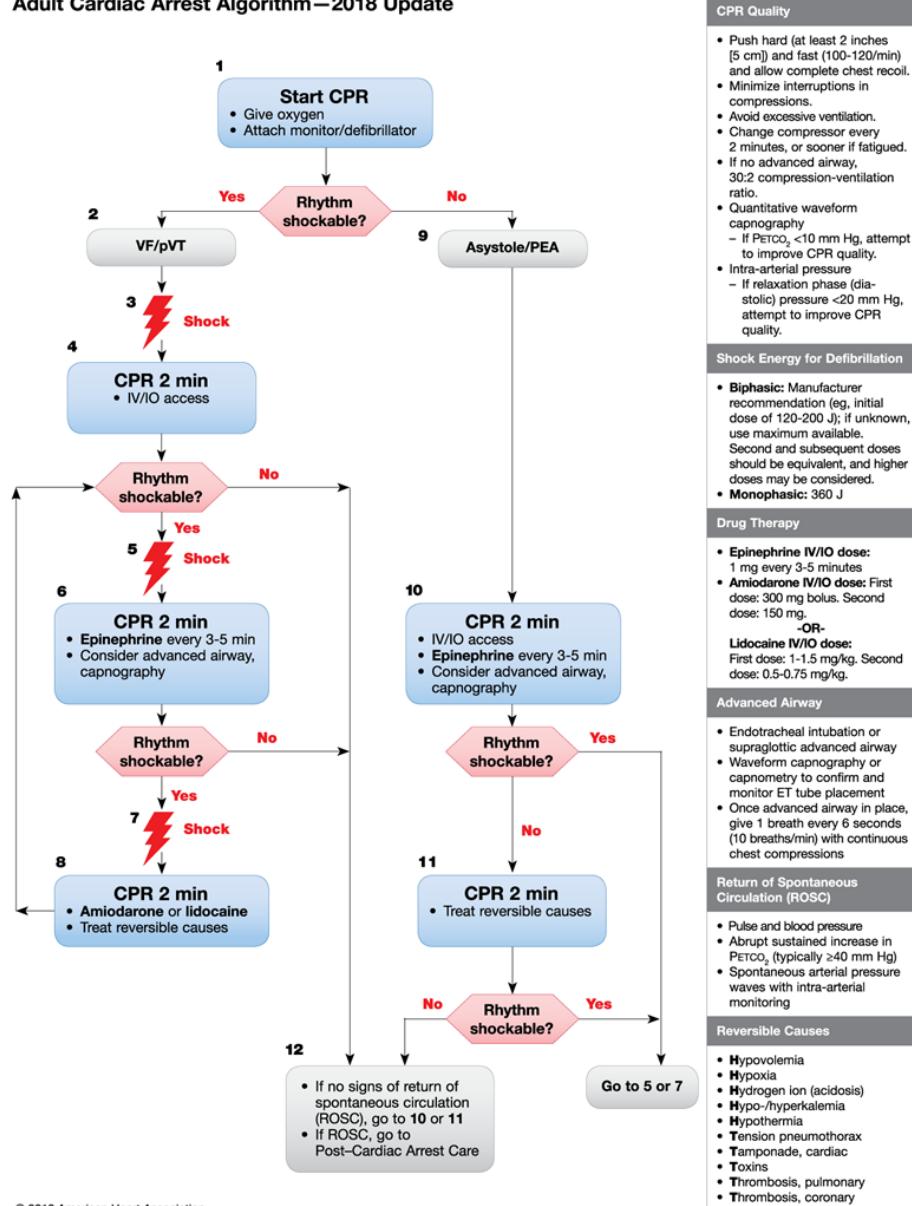
monitored by a pad and feedback is given to the user to maintain within the guideline recommendations. Compressions are applied to a chest pad that is between the hand and the chest and an accelerometer in the pad is used to measure the compression depth and rate. This information is fed back to user via the display and voice prompts. This technique can only be used when the patient is on a rigid surface.

A logic next step was the introduction of mechanical devices which offered guaranteed compressions as long as there is a power source. In recent large studies the outcome of cardiac arrest events was not improved when the mechanical devices were compared with manual CPR. There is at present no recommendation to use these devices.

15.7.1 Guidelines for Cardiac Arrest Treatment

The American heart association (AHA) and the European Resuscitation Council (ERC) provide new guidelines for resuscitation care every 5 years. The recommendations are based on a thorough review of the literature and experienced teams formulate recommendations for clinical care during resuscitations. A difference is made for basic life support (BLS for lay persons, less experienced EMS, clinicians) and advanced life support (ALS) for highly trained and well equipped medical teams or EMS personnel. The 2018 changes to the 2015 ALS guidelines are shown in Figure 15-23. The steps described in this figure are quite general and are characterized as one size fits all. They are aimed for the large average patient group. This approach is also beneficial for training. Chest compression depth and frequency are not differentiated for gender, size and weight of the patient. The compressions depth target is 5-6 cm, this may be too much for smaller patients and too small for large patients. This depth was selected using data from clinical studies with QCPR type functionality. This depth is set to avoid shallow compressions in the majority of patients, too small depth strongly degrades outcome. There is a need for more personalized treatment by optimizing the compression parameters to a specific patient. However monitoring techniques that can be used during CPR are scarce and evidence for improved outcome needs to be generated. In the next two sections the statistics of cardiac arrest and results from a laboratory test are described. The results form a laboratory study are shown to give insight in physiologic parameters during CPR and the possible use of advanced monitoring techniques during CPR to personalize care and improve outcomes.

Adult Cardiac Arrest Algorithm—2018 Update



© 2018 American Heart Association

Figure 15-23 ERC 2018 guidelines for advanced life support.

15.7.2 Statistics of sudden cardiac arrest

Cardiac arrest is a leading cause of death in the western world and Asia. In 2018 more than 356000 people suffered an out of hospital cardiac arrest (OHCA) in the United States, the survival to hospital discharge was 10.8%, only 9 % of all victims had a good neurological function. In the USA there are large regional differences in survival rate, it varies between 3.4% and 24%. The survival rate depends most on the time to treatment, quality of resuscitation care and training of the EMS organization and on the fraction of bystander CPR.

The district of King-County near Seattle has one of the best EMS systems in the world and outcomes of OHCA are among the best in the world. Data from 2016 are shown in Figure 15-24. This district has an extremely well organized EMS system. There are multiply tiers (lay responders, police, fire brigade, EMS) to reduce response time. Bystander response is high (73% of cases had bystander

CPR). Police or fire brigade is the first tier and average response time is within 6 minutes followed by the arrival of a specialized EMS unit within minutes. This short time to treatment and top quality resuscitation care by all stakeholders is an important factor in the high survival rates.

VT/VF was the dominant rhythm in about 25% of the cases and the average out of hospital survival rate was 52% for this group. PEA was observed in 31% of the events and survival was 24%. Asystole was the most frequent rhythm and was observed in 41% of the cases, the survival rate was 4%. When the arrest occurred after arrival of the EMS survival rates were much higher for PEA and asystole.

Only 20 to 26% of cardiac arrests was due to a shockable rhythm where the use of an AED make sense. For VF/ VT and even more for the other rhythms high quality CPR and other resuscitation care are of paramount importance for survival. In the following section data obtained from a laboratory test of a high quality mechanical CPR device are shown to give insight in physiologic parameters and the values that are important for survival.

	Number Treated	Number Survived To Hospital Discharge	Percent Survived
Arrest Before Arrival of EMS	1,086	233	22%
Ventricular Fibrillation/ Tachycardia (VF/VT)	284	143	50%
Asystole	472	15	3%
PEA	303	66	22%
Not Shockable, but unknown if PEA or asystole	14	1	7%
Unknown	13	8	62%
Arrest After Arrival of EMS	142	55	39%
Ventricular Fibrillation/Tachycardia (VF/VT)	34	22	65%
Asystole	24	6	25%
PEA	79	27	34%
Not Shockable, but unknown if PEA or asystole	1	0	0%
Unknown	4	0	0%
Total	1,228	288	24%

Figure 15-24 Summary outcome data of OHCA in 2016 in King County USA

15.7.3 Monitoring during CPR – A case study in the laboratory

Monitoring during CPR may be used to optimize and personalize resuscitation treatment and improve outcome. Monitor data obtained during a laboratory study of a prototype mechanical CPR device are described and interpreted. Guideline based CPR with 30 compressions followed by a short pause for two manual ventilations (30:2) was used. The protocol is shown schematically in Figure 15-25.

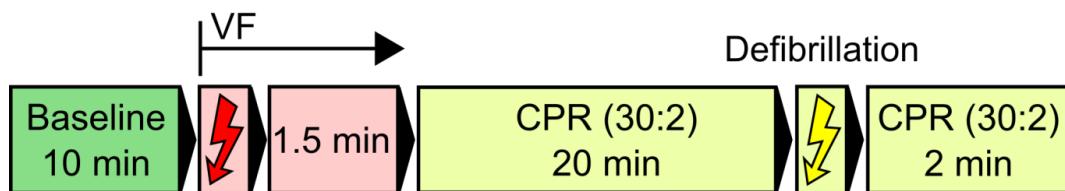


Figure 15-25 Summary of the animal laboratory protocol.

The preparation and CPR parameters were optimized for the specific animal model. Aortic and right atrium blood pressure, capnography and ECG data are shown in Figure 15-26 for the base line (10 minutes), CPR (20 minutes) and ROSC periods. The first 600 seconds are the baseline period for the reference data extraction. After 600 seconds VF is induced by an electric pulse and the heart rhythm

of VF was identified. Blood pressure dropped immediately and partial pressure of expired CO₂ decreased exponentially with every ventilation. At 690 seconds mechanical CPR was started. The compression depth was 4 cm and the compression frequency was 100 per minute. The effective compression depth of 4 cm was maintained during the CPR period of 20 minutes. The CO₂ pressure in expired air increased slowly and reached acceptable values. The increase with time and the magnitude of ETCO₂ confirmed that CPR quality was good (there is blood flow and metabolism). The peak arterial blood pressure decreased significantly with respect to base line and the peak atrial blood pressure was similar in magnitude to that of the aortic pressure.

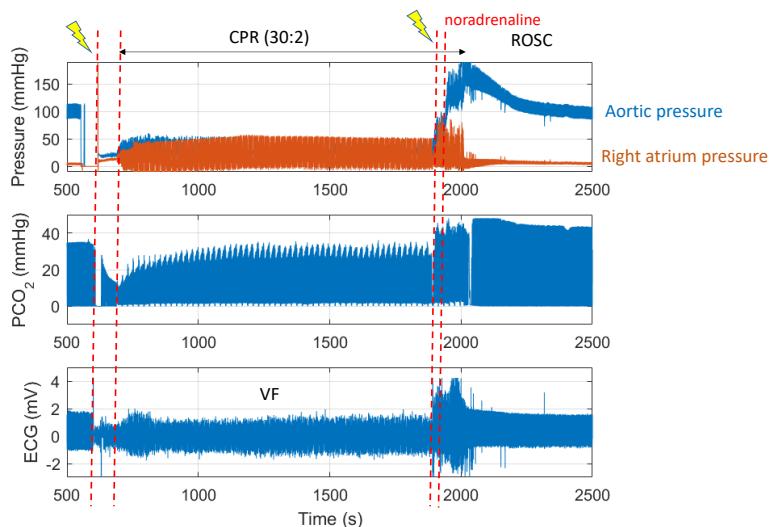


Figure 15-26 Data of animal test of mechanical CPR device during baseline, 30:2 CPR and ROSC periods.

The carotid blood flow is shown in more detail in Figure 15-27. During base line period and ROSC periods flow is always in the forward direction and mean flow is around 300 ml per minute or larger. During CPR, flow has both forward and reverse components. The peak forward flow is large but the large reverse flow reduces net flow. The mean or net forward flow is 20 to 25 percent of the base line flow during CPR.

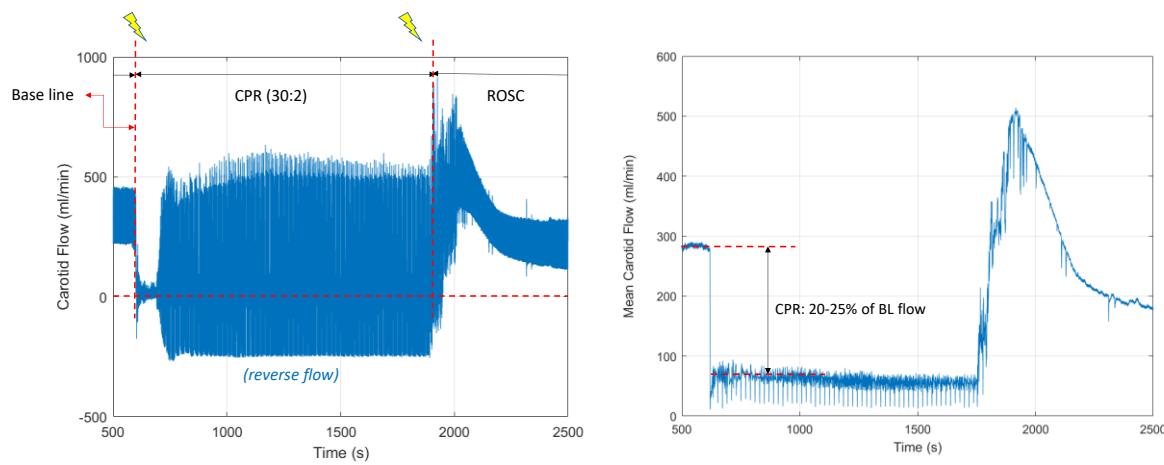


Figure 15-27 Carotid blood flow (right) and mean carotid blood flow during base line, 30:2 CPR and ROSC period.

The arterial and right atrial blood pressures during CPR are shown in more detail in Figure 15-28. The maximum systolic pressures are approximately 50 mmHg which is much smaller than the blood pressure in the base-line period. During the recoil phase of the compression the aortic pressure remains at 30 mmHg while that of the right atrium drops to 4 mmHg. The difference between two

pressures is the coronary perfusion pressure (CPP). The CPP parameter is the driving pressure for coronary blood flow and should be large enough for sufficient perfusion of the heart muscle. The CPP parameter measured during CPR has been found to be the best indicator for ROSC. A value of 15 to 20 mmHg is a threshold for ROSC. A larger CPP is better. In this case CPP is approximately 25 mmHg which is above the threshold value.

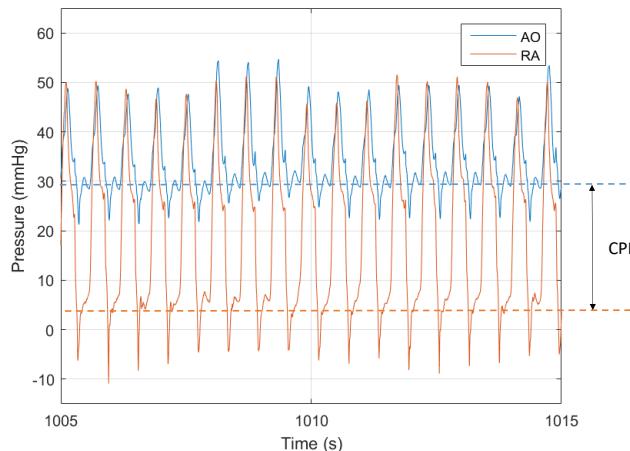


Figure 15-28 Aortic (AO) and right atrial (RA) blood pressure waveforms during CPR. The dashed lines are estimates of the diastolic pressures in the aorta and right atrium during the recoil phase. CPP is the coronary perfusion pressure, i.e. the arterial-venous pressure gradient over the coronary blood vessels.

During the no-flow period after VF induction and the CPR period (total 20 minutes) the arterial pressures and carotid flow remained stable and were at good levels. After a CPR period of 20 minutes a defibrillation shock was applied and the VF rhythm was converted to a normal perfusing rhythm and increases in blood pressure was detected. It was found that blood pressure were not large enough and CPR compressions continued (see for instance Figure 15-30). The decision to stop CPR is difficult, it cannot be based on the ECG alone as the heart rhythm may not be perfusing (PEA). It is known that CPR on a beating heart can induce VF, therefore compressions should be stopped when ROSC is stable. In some cases severe injury of the cardio-vascular and pulmonary systems has been observed. Stopping too early may revert the ECG rhythm from perfusing to VF again. There is a need for measurements that support the clinician in the decision to stop or continue CPR. CPR is stopped when the clinician detects a stable and perfusing rhythm (the pauses are ideal for this process). Just after the shock blood pressure was too low and heart rate was not stable. It was deemed necessary to inject noradrenaline to strengthen the heart contractions. Blood pressure and flow increased abruptly, the heart was now perfused well, the situation improved and it was decided that the perfusing rhythm was strong and stable and CPR was stopped.

Laboratory and clinical studies have shown that there are physiologic parameters that correlate with outcome. It has been proposed to use monitoring techniques during CPR and use this feedback to optimize CPR. It appeared that CPP is the most important parameter for ROSC and survival, it is linked to the perfusion of the heart. Two invasive blood pressure measurements are needed and this is not feasible or practical during a real CPR event. Peak pressures in blood pressure are less relevant. Non-invasive methods to measure blood pressure can only be used in rare cases during CPR. Another parameter that is important for neurological outcome is the net flow of oxygenated blood to the brain, a value of 25% of baseline flow is on the low side a larger net flow would be better. In the present case flow was measured with an ultrasonic transit time transducer which requires surgical techniques for placement, this cannot be done during a real event. An external ultra sound probe could be used but is also unpractical during CPR, a skilled operator is needed and this may interfere

with the care process. Furthermore motion artifacts could corrupt the real flow related data. Small disposable continuous-wave Doppler transducers have been proposed. This is an option when motion artifacts can be suppressed and when such devices become available at an affordable cost.

Photo plethysmography (ppg, pleth) and ECG are non-invasive techniques and are available during CPR when a monitor-defibrillator is present. ECG is used during CPR for rhythm analysis. ECG and plethysmography signals measured during CPR and ventilation pauses are shown in the left side of Figure 15-29. The ECG signal is affected by motion and during pauses the VF signal is observed. There was no pulsatile ppg signal during pauses as the heart does not have a perfusing rhythm.

Signals measured during the ROSC period are shown in the figure on the right side. The time interval includes the moment when it was decided to stop CPR. An ECG signal that corresponds with a perfusing rhythm was already visible after the defibrillation shock. However this ECG signal alone does not prove that there is a perfusing rhythm, PEA occurs frequently after defibrillation. Note that the transition between compressions and no compressions is clearly visible in both signals. During the CPR period in this figure the ppg signal consists of physiologic pulsatile signals and CPR induced artifacts. These artifacts are absent when there is no CPR. It appeared that already very small physiologic ppg signals due to ROSC and a beating heart could be detected during CPR. There is a difference in frequency between the spontaneous heart beat and compression related signal. These findings indicate that the combination of the ECG signals and a pulsatile ppg could be used to detect ROSC during CPR and certainly during compression pauses.

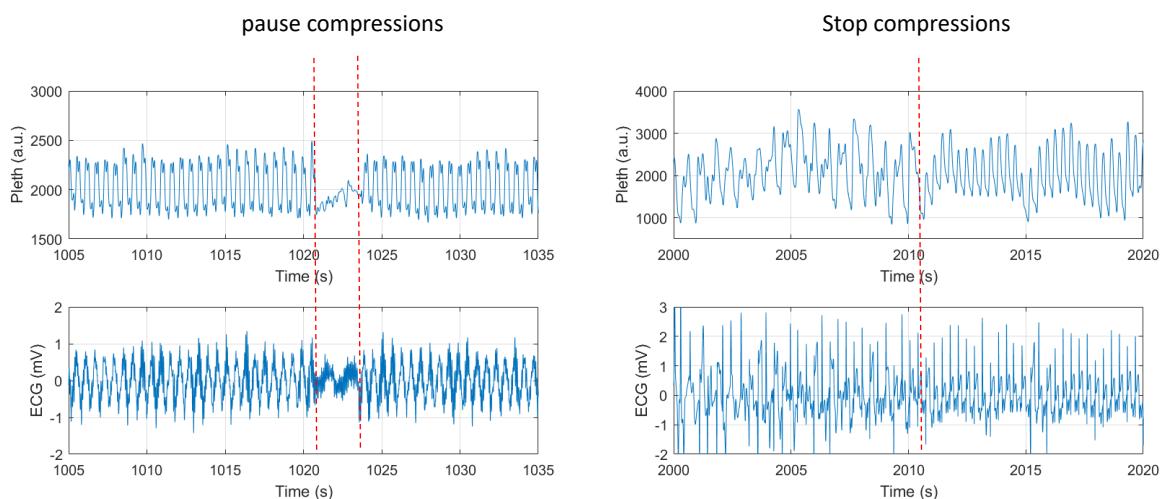


Figure 15-29 Photo plethysmography (Pleth, ppg) signal and ECG during CPR and pauses (left) and at ROSC with/without compressions.

The picture at the right side shows that a perfusing heart beat changes the ppg signal during CPR. In principle the ppg signal could be separated from the motion artifact when the compression is recorded. It was decided to study this method of reliable ROSC detection during CPR in more detail by a more detailed analysis of the data and in new experiments. Results of this follow-up study by R. Wijshoff (published in Resuscitation 2013) are shown in Figure 15-30. The ac-component of the IR ppg sensor that was mounted on the nose of the animal and the arterial pressure are shown in the period before and after the defibrillation shock. Before the shock no pulsations are observed in blood pressure and ppg during the pauses. After the shock new pulsations in both the arterial pressure and ppg signals are observed during CPR and during pauses. A pulsatile flow is present but arterial pressure is still too low to stop CPR. The heart has to become stronger before CPR can be stopped.

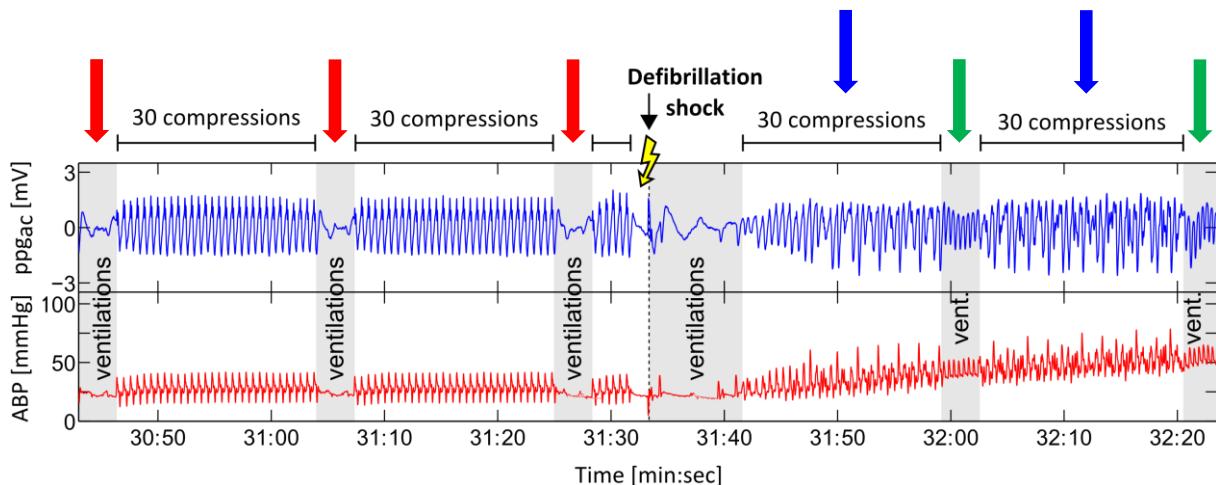


Figure 15-30 AC amplitude of the IR ppg signal and arterial invasive blood pressure before and just after the defibrillation shock. (R. Wijshoff et al, Resuscitation 2013).

The frequency spectrum of ppg was studied and the spectrogram is shown in Figure 15-31 for the time interval before and after defibrillation. Before the shock the ppg signal only has frequency components that correspond with the compression frequency. After the shock the compression component is still available but a second component due to the beating and perfusing heart can be observed. Initially the heart rate increases and decreases, after administration of adrenaline the spontaneous heart beat becomes stronger, ppg amplitude and blood pressure increase and CPR was stopped. The ppg signal does not provide information on blood pressure which is actually needed but it is suited for decision support. A clinician could measure the blood pressure using a non-invasive method to validate the decision. This ROSC detection method during CPR has been studied for manual CPR and is presently further validated in clinical studies. Motion artifacts limit the use of the photoplethysmography signal for optimization of compressions during CPR.

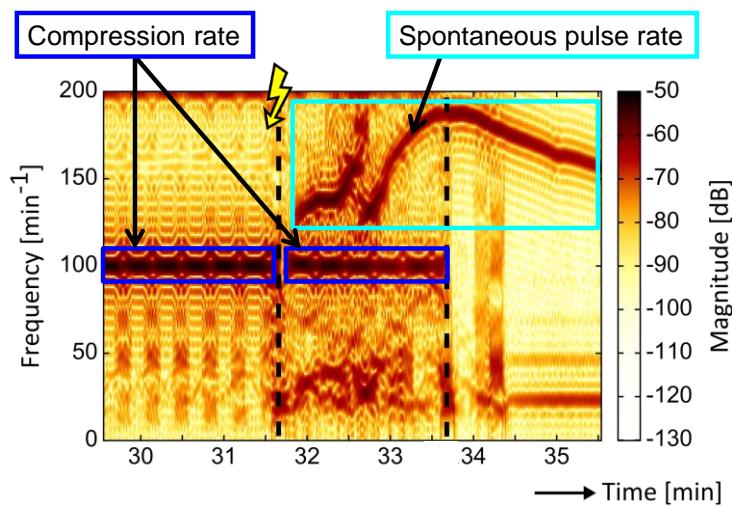


Figure 15-31 Frequency spectrogram of the ppg signal for the same time period.

When capnography is available it recommended in the guidelines to be used for optimization of CPR quality. The measurement is available in most monitor-defibrillators. End-tidal CO₂ data are shown in Figure 15-32. After VF induction there is no blood flow and supply of new CO₂ gas to the capillaries in the lung and ETCO₂ drops rapidly. With each ventilation the CO₂ pressure drops with the same

fraction. When CPR is started, ETCO₂ increases sharply. CPR induced blood flow perfuses the lungs with CO₂ rich blood and the gas exchange increases. During CPR ETCO₂ increases slightly, this is mainly caused by increase of the CO₂ pressure in blood (note that the cardiac output has dropped with a factor 4 to 5 and anaerobic metabolism occurs). At ROSC blood flow increases, the CO₂ concentration in the capillaries rises and ETCO₂ shows a sharp rise. The ETCO₂ pressure is initially very large due to the high CO₂ gas pressure in blood. After administration of adrenaline cardiac output increases strongly and ETCO₂ rises again.

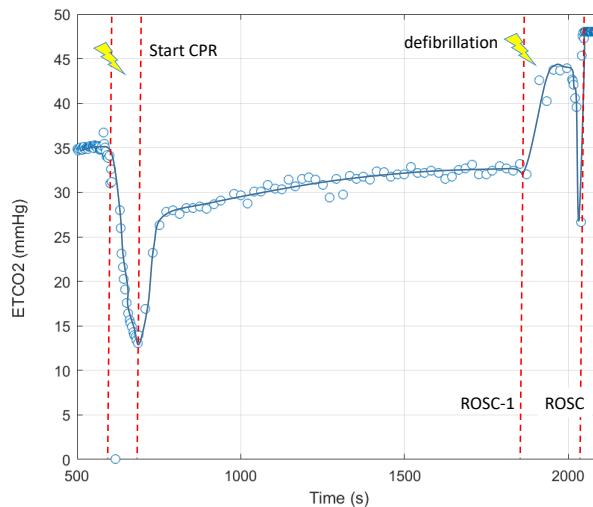
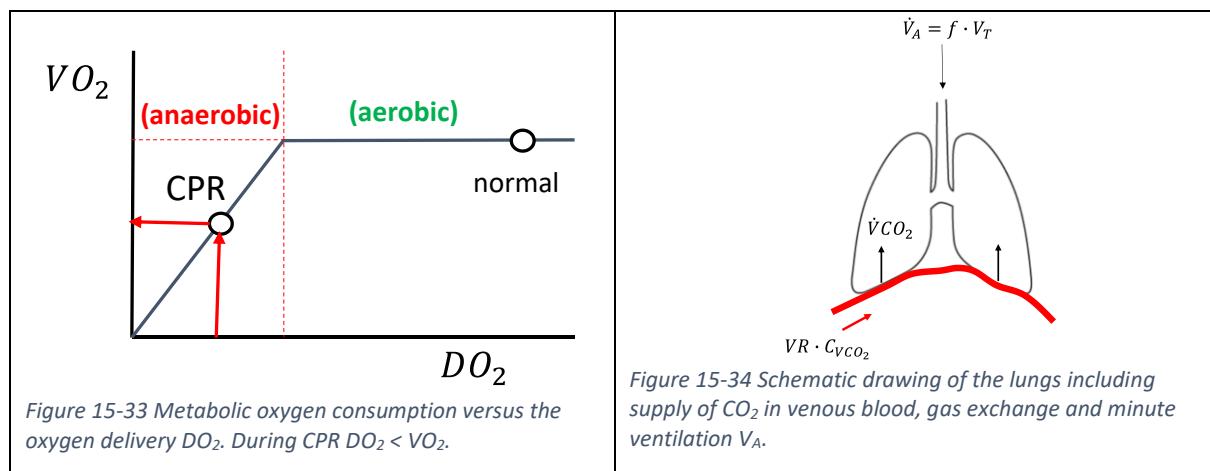


Figure 15-32 ETCO₂ data during animal test. VF induction, start CPR, defibrillation, first ROSC and final ROSC are indicated in the figure.

Capnography gives information on the ventilation status and gives indirect information on the circulation and metabolism. During CPR the oxygen delivery by arterial blood is smaller than the metabolic need in rest (see Figure 15-33). The cells switch over to the anaerobic metabolism and lactic acid is released, the pH decreases and PCO₂ in tissues and capillary blood increases. The oxygen pressure in the tissue and capillary blood decreases.



In this case venous blood will have both a lower oxygen saturation and a higher partial pressure of carbon dioxide. The CO₂ rich blood flows to the lung where gas exchange between the lung and capillary blood takes place. The transport rate of CO₂ from blood to the alveolar space is given by Fick's equation:

	$\dot{V}_{CO_2} = D_L \cdot (P_{VCO_2} - P_{ACO_2})$	15.7-1
--	--	--------

P_{VCO_2} is the partial pressure of CO₂ in venous blood and P_{ACO_2} is the partial pressure of CO₂ in the alveolar space. D_L is the lung diffusion capacity. The CO₂ partial pressure in the lung is given by:

	$P_{ACO_2} = \frac{\dot{V}_{CO_2}}{\dot{V}_A}$	15.7-2
--	--	--------

\dot{V}_A is the minute ventilation of the alveolar space. The venous CO₂ delivery (DCO_2) is equal to:

	$DCO_2 = VR \cdot C_{VCO_2}$	15.7-3
--	------------------------------	--------

VR is the venous return and C_{VCO_2} is the CO₂ concentration in venous blood. When the venous return is zero, there is no supply of new venous blood and the CO₂ partial pressure in venous blood and lung will decrease after each ventilation until the partial CO₂ pressure in blood and lung are negligibly small. This is observed during the no flow time in Figure 15-32. When CPR is started there is a venous return blood flow, blood with a high PCO₂ is supplied to the lung and CO₂ diffuses to the alveolar space and is partially exhaled. The pressure of CO₂ in the lung depends both on the supply rate DCO₂, the CO₂ capillary-alveolar pressure gradient and the minute alveolar ventilation \dot{V}_A . The ETCO₂ pressure is an indicator of the quality of lung ventilation, is proportional to the venous return blood flow and shows that there is metabolism in the tissues (O₂ consumption and CO₂ production). The minute volume (both tidal volume and ventilation rate) have to be controlled well when capnography is used during CPR optimization, the value depends on the minute ventilation. Note that in practice both the magnitude and trends in ETCO₂ are of importance. A low (< 10 mmHg) and/or steady decreasing ETCO₂ value are indicators of poor outcome. The rescuer may react on this trend by applying more forceful and faster chest compressions in the hope that this increases CPR induced blood flow. At present capnography is the method that is used most often to tune and adjust CPR during resuscitations. New monitoring techniques that give information of blood flow to the brain and/or on the coronary perfusion pressure are needed.

15.8 Image Guided Therapy

The combination of imaging, monitoring and therapy is called image guide therapy. The number of use cases is growing rapidly. Coronary angioplasty is by now the main technique to treat patients with an acute myocardial infarction. The method is illustrated in Figure 15-35.

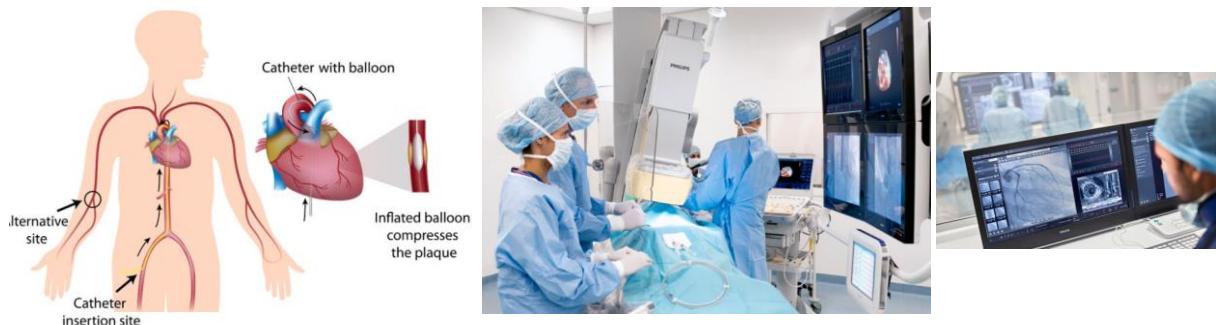


Figure 15-35 Illustration of an image guided therapy in the Cath lab.

After a suspected blockage of one or more coronary arteries a patient is transported to the Cath lab for investigation, diagnosis and treatment when deemed necessary. The patient is placed on a table of a high resolution X-ray imaging system. Monitor data obtained from the catheter and patient monitor are shown on a large screen and on a second screen high-resolution images are displayed at video rates. The cardiologist inserts a special catheter at the femoral site and moves the catheter via the aorta to site where the large coronary arteries receive blood from the aorta. Then a contrast agent is released and the clinicians and an operator in the control room can see in real time which coronary artery is affected and can determine the location of the stenosis in the artery. From the image the degree of blockage can be estimated, this estimate is crude and in some cases the diagnosis if a stent is needed or not can be incorrect.

In advanced monitoring systems the catheter is equipped with multiple high-fidelity (MEMS) pressure sensors. The catheter is further moved into the coronary vessel and blood pressure waveform is measured before and after the stenosis. The iFR ratio is determined. “The instantaneous wave-free ratio (iFR, sometimes referred to as the instant wave-free ratio or instant flow reserve) is a diagnostic tool used to assess whether a stenosis is causing a limitation of blood flow in coronary arteries with subsequent ischemia.” The iFR is the ratio of blood pressures determined proximal and distal of a stenosis. The part of the diastolic phase of the pressure waveform is used where reflections are absent. It was found an iFR ratio below 0.9 is an indication of flow limitation. For a low iFR a stent is placed at the site of the stenosis and the flow limitation is removed. This can be validated by measuring iFR again. An even more advanced method is the use of IVUS (Intra Vascular Ultra Sound). In IVUS a small ultra sound transducer is placed at the tip of the catheter and degree of blockage can be quantified after image processing. The three methods to determine the degree of blocking at a stenosis are compared in Figure 15-36.

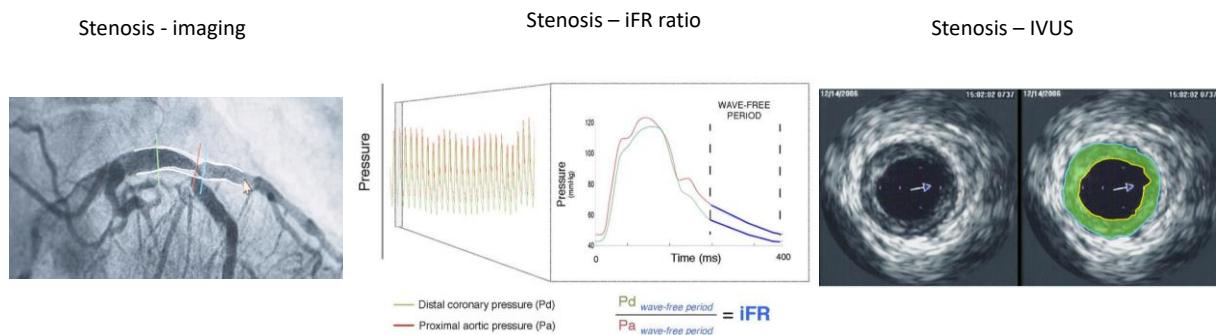


Figure 15-36 Methods to determine the degree of blocking at a stenosis site. Left imaging, middle iFR ratio from the blood pressure waveform and right by imaging with IVUS. (Blue line is coronary artery diameter, the yellow line is the lumen diameter)

X-ray imaging was the dominant method for quantification of the degree of flow reduction but specificity is not high enough. The blood pressure IFR method and the IVUS method are much better indicators of the degree of flow limitation of blood flow and are now becoming more popular. The image guide therapy is become more used for more and more applications. For example it is now used for replacement of the heart valves.

15.9 Pacemakers [8]

"A pacemaker (or artificial pacemaker) is a medical device that generates electrical impulses delivered by electrodes to contract the heart muscles and regulate the electrical conduction system of the heart." (See Figure 15-37). This device is needed when the heart's natural pacemaker is not fast enough, there is a block in the heart's electrical conductive system, asynchronous ventricle contraction. The cause of electrical heart failure is often due to problems in the generation or conduction of depolarization in the heart muscle. The AV-node in particular is vulnerable. It consists of a bundle of very small nerve fibers. It is located between atrium and ventricle, and its function is to delay the depolarization and contraction of the ventricle until it has been adequately filled by the atrium. If its function is lost, the ventricle still contracts but at a far too low rate. This condition is called *heart block*. An implanted pacemaker can resolve this problem in heart block patients. Before the implant operation, an external pacemaker can keep the patient in a fit condition.

A pacemaker can be programmed (in an implanted one non-invasively, through a "magnet" on the skin) into a variety of "pacing modes". In the simplest of these, a small electrical pulse (0.1 to several mA) with a width of 0.5 to 2 ms is applied to the ventricle at a fixed rate of about 60 pulses per minute through a flexible, catheter-like conducting electrode that runs from pacemaker to ventricle, usually through a vein ("asynchronous pacing"). In a more physiological mode, one electrode affixed to the atrium detects the atrium's depolarization, which is recognizable as the P-wave of the ECG; after an appropriate delay during which the ventricle is filled with blood, the pacemaker delivers a pulse to the ventricle ("atrium-synchronous pacing"). If the heart block is not total, normal conduction may be disrupted only occasionally. In such cases, the pacemaker is programmed into a mode, which applies a pulse to the ventricle only if the ventricle's depolarization does not occur within a certain time after the atrium's depolarization has been detected ("demand pacing").

Knowing whether a patient has an implanted pacemaker is important, because defibrillation may damage or reprogram it. In elective cases, the patient's record will show this information (if it can be found!); in emergencies, records are often unavailable. The ECG will often show whether a pacemaker is present. Sometimes, pacemaker pulses are directly visible in the ECG. Pacing pulses are so short and their energy is low, however, that they are frequently invisible on the display of standard ECG-monitors, whose bandwidth is limited. Yet, even then a more detailed analysis of the

ECG can often show the presence of a pacemaker. In asynchronous pacing, the ECG derived heart rate, which is based on the time between successive QRS-complexes, is utterly constant, whereas a non-paced ECG always shows some irregularities. In atrium-synchronous pacing, likewise, the time between a P-wave and the following QRS-complex is more constant than normal. In demand pacing, however, the presence of a pacemaker cannot be discovered as long as the patient's ventricle functions normally; then the pacemaker is "silent".

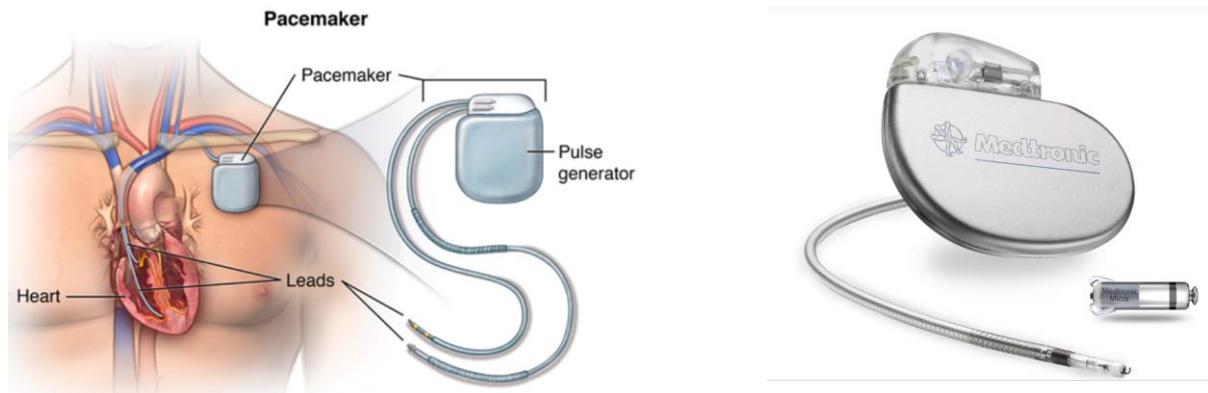


Figure 15-37 Schematic drawing of a pacemaker implanted in the chest of a patient. Lead wires are connected via a vein or artery at specific locations in the heart.

15.10 Heart Lung Machines [8]

Heart surgery is often delicate and possible only if the heart is made motionless. In that case, the pumping function of the heart is lost and a heart-lung machine must ensure the continuation of circulation; this is called *extracorporeal circulation* (ECC). The gas exchange function of the lung is taken over as well, because technically it is simpler and safer to realize the functions of both the left and the right heart with one pump. This also eliminates the thorax movements that are normally caused by ventilation. Figure 15-38 shows a diagram of a heart-lung machine.

The main *pump* is the artificial heart. It pumps a flow of blood (the artificial "cardiac output") into the aorta; blood flow is proportional to the pump's angular velocity. Usually an arterial pressure of about 70 mm Hg is maintained. In contrast to a normal arterial pressure, it is now non-pulsatile. Thus the "systolic" (maximum) and "diastolic" (minimum) value do not differ from the mean value. When a roller pump (see section 8.1) is used, the blood comes into contact with a flexible nylon or polyester hose only. These materials limit damage to the blood as much as possible. Centrifugal pumps result in even less blood damage and can thus be used for longer time periods.

The *oxygenator* (gas exchanger) is the artificial lung. Normally, a membrane oxygenator is used. Blood flows on one side of the membrane; on the other side, a liquid circulates whose O₂ and CO₂ partial pressures are controlled in such a way that desired blood gas values result. This is done by bubbling adjustable flows of O₂, CO₂ and air through the liquid. Usually these flows are adjusted by and read off from a *rotameter block*. The advantages of membrane oxygenators are the avoidance of gas emboli (these are formed in a bubble oxygenator in which gas is bubbled through the blood) and minimal damage to the blood, of which especially the platelets (thrombocytes) are easily disrupted. Disruption of platelets, which causes coagulation (clotting) of the blood, cannot be prevented completely. Coagulation is attacked by adding large amounts of heparin (an anticoagulation agent) to the blood.

Additional pumps force leakage blood and lung venous blood that may have accumulated back into the circulating volume. Filters remove blood clots and air bubbles from the circulation. Flexible hoses connect the parts of the heart-lung machine together and thick cannulas connect the hoses to the

patient's (inferior and superior) vena cava and aorta. The venous outflow is passive; it is due to the hydrostatic pressure difference that results from the lower position of the heart-lung machine's blood reservoir. A heat exchanger controls the blood temperature. The artificial heart forces the blood through the oxygenator and, after gas bubbles and blood clots have been removed, into the aorta. The heart-lung machine adds about 2 l to the circulating blood volume in an adult, half of this in children and babies. This extra volume is (donor) blood, blood plasma or plasma similar fluids (colloidal electrolytes).

Hypothermia (a reduction of the patient's temperature) to about 28 °C reduces the tissues oxygen consumption and carbon dioxide production with about 50%. This provides a safety margin in this artificial situation, where perfusion is far from normal. The heat exchanger consists of a heat exchange surface, usually of a metal coated with nylon (polyamide) or a similar material. Blood flows on one side. On the other side flows water, of which the temperature can be controlled. At the start of the operation, cold water cools the patient; at its end, warm water reheatsthe patient to 34°C. Most operations can be done at moderate hypothermia (down to 25°C). In some operations, the blood flow must be stopped completely. Since metabolism is reduced to about 50% at 27°C, about 6 minutes at most are then available for surgery. Deep hypothermia at 19°C extends this period to 30 minutes.

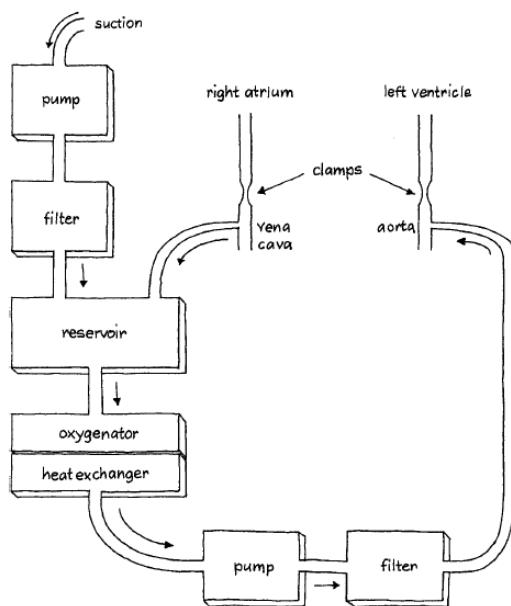


Figure 15-38 Schematic diagram of a heart-lung machine.

Modern heart-lung machines have a computer interface through which they can make their internal settings and measurements (flows, pressures, temperatures) available to the outside world. These data can be processed and used to automatically generate reports, "intelligent" alarms, and to automate certain functions.

15.11 References

1. Otis, A. B., C. B. McKerrow, R. A. Bartlett, J. Mead, M. B. McIlroy, N. J. Silverstone and E. P. Radford (1956). Mechanical factors in distribution of pulmonary ventilation. *J. Appl. Physiol.* 8: 427-443.
2. Mead, J. and J. L. Whittenberger (1954). Evaluation of airway interruption technique as a method for measuring pulmonary air-flow resistance. *J. Appl. Physiol.* 6: 418-416.
3. Iotti GA, Braschi A, Brunner JX, Smits T, Olivei M, Palo A, Veronesi R (1995) Respiratory mechanics by least squares fitting in mechanically ventilated patients: applications during paralysis and during pressure support ventilation. *Intensive Care Med* 21: 406-413.
4. Arnal, Schaedler and Kiralki in G. Bellani (ed.), Mechanical Ventilation from Pathophysiology to Clinical Evidence, Chapter 12 pp 127-138, https://doi.org/10.1007/978-3-030-93401-9_12.
5. ARDSnet. *N Engl J Med.* 2000 May;342(18):1301-1308.
6. M. Botta, E.F.E. Wenstedt, A.M. Tsonas, L.A. Buiteman-Kruizinga, D.M.P. van Meenen, H.H.M. Korsten, J. Horn, F. Paulus, A.G.J.H. Bindels, M.J. Schultz and A.J.R. De Bie, Effectiveness, safety and efficacy of INTELLiVENT—adaptive support ventilation, a closed-loop ventilation mode for use in ICU patients – a systematic review, *EXPERT REVIEW OF RESPIRATORY MEDICINE*, 2021, VOL. 15, NO. 11, 1403–1413, <https://doi.org/10.1080/17476348.2021.1933450>.
7. Neurally Adjusted Ventilatory Assist in Acute Respiratory Failure—A Narrative Review, Michele Umbrello , Edoardo Antonucci and Stefano Muttini, *J. Clin. Med.* 2022, 11, 1863. (<https://doi.org/10.3390/jcm11071863>).
8. J.A.Bлом, Monitoring of respiration and circulation, EAN 9780849320835

16 ANSWERS TO QUESTIONS

16.1 Answers Chapter 3

- 1) Pressure is a scalar quantity and is defined as the force per unit area, a force is a vector quantity, it has both magnitude and direction.
- 2) The internal pressure is the pressure exerted on a wall due to momentum transfer by colliding particles (atoms, molecules) in the fluid. The dynamic pressure is the pressure component due to collective motion of the particles in the fluid, it is equal to $\frac{1}{2} \rho v^2$, ρ is the fluid density. The hydrostatic pressure is the pressure in a certain point in the fluid due to gravitational forces, exerted by the weight of the fluid above a certain point, it is equal to $\rho g(h - h_0)$, where h is the height difference between the point and a reference point h_0 , and g is the acceleration in gravity ($g=9.807 \text{ m/s}^2$).
- 3) SI unit of pressure is N/m^2 or Pascal. One atmosphere pressure is equal to 101325 Pa. One atmosphere is also equal to 760 mmHg. Therefore $1 \text{ mmHg} = \frac{101325}{760} = 133,3 \text{ Pa}$. The density of water at 20 °C is 998.19 kg/m^3 . The height of a column of water in a manometer is then $\frac{101325}{998.19*9.807} = 10,35 \text{ m} = 1035 \text{ cm}$. Therefore $1\text{cmH}_2\text{O}$ at 20 °C is equal to $101325/1035 = 97,9 \text{ Pa}$.
- 4) After elastic deformation when the force is removed the object returns to its original shape and dimensions, changes are reversible. After plastic deformation the shape and dimensions are changed, the deformation is irreversible.
- 5) Springs in parallel, have same elongation Δx , forces add $F = F_1 + F_2 = k_1 \cdot \Delta x + k_2 \cdot \Delta x = (k_1 + k_2) \cdot \Delta x$, the spring constant k is $k = k_1 + k_2$. Springs in series, the force is the equal in both springs, the elongation is the sum of two elongations, $\Delta x = \Delta x_1 + \Delta x_2 = \frac{F}{k_1} + \frac{F}{k_2} = \frac{F}{k}$. Hence the force constant of the combined spring is $\frac{1}{k} = \frac{1}{k_1} + \frac{1}{k_2}$ or $k = \frac{k_1 k_2}{k_1 + k_2}$.
- 6) The Poisson the negative of the ratio of transverse strain and axial strain.
- 7) In steady state the forces due to pressure and elasticity balance each other. Split the sphere in two equal halves and consider the two force components perpendicular to the cross sectional plane. The elastic force is equal to: $F_{el} = \sigma * A = \sigma * d * 2\pi r$. A is the area of the thin wall cross section, d is the wall thickness, σ is the wall stress and r is the radius of the sphere. The force due to the pressure is equal to: $F_p = \pi r^2 \Delta P$. In equilibrium the two forces balance each other and it follows that: $\Delta P = \frac{2\sigma d}{r} = \frac{2T}{r}$. T is the surface tension. The wall tension is two times smaller than in a cylinder, all other parameters being the same.
- 8) From the Laplace equation it follows that the pressure in the balloon with small radius is larger than the pressure in the balloon with a large radius. When the valve is opened air will flow from high pressure to low pressure, i.e. from the small radius balloon to the larger radius balloon.
- 9) Stress is the force per unit area, strain is the relative change in dimension due to a stress. The Young modulus is the material dependent proportionality constant between stress and strain in the elastic deformation range. A large E corresponds with a stiff material.
- 10) The stress-strain relation of biologic tissue is non-linear. A plot of stress on the y-axis and strain on the x-axis shows a j-shaped curve, exponential curve. Initially the material is very elastic, the incremental Young Modulus is small. For larger strain the stress increases exponentially, the material is stiff. In summary, for small strain it deforms easily, for larger strain it is much stiffer.

- 11) The elastic compliance C is defined as: $C = \left(\frac{dV}{dP}\right)_P$ the incremental change in volume for a small change in pressure at a transmural pressure P . A biologic material deforms easily at low transmural pressure, hence compliance is large. At larger transmural pressure the material is much stiffer and the compliance is much smaller.
- 12) In a visco-elastic material both elastic and viscous forces are present. Hence these materials have properties of a solid and of a fluid, due to the viscosity the material cannot deform instantaneously. In an elastic material the viscous damping forces are not present. The viscous forces cause energy dissipation in the form of heat, in elastic materials energy is stored in the form of potential elastic energy and energy is conserved.
- 13) Hysteresis is observed in the stress-strain relation of viscoelastic materials. The stress-strain curve for during increasing strain differs from the curve during decreasing strain. The area of the loop is proportional to the dissipated energy.
- 14) Biologic visco-elastic materials are very complex and often highly non-linear. The simplest model of visco-elastic materials have two lumped elements, one spring and one damper. The series combination is the Maxwell model, the force is the same in both elements. When a sudden fixed strain is applied to this series combination, initially the total strain is equal to the strain the spring $\Delta l = F/k$, thereafter the damper starts to move and the spring strain decreases until the strain is due to the damper only, the stress and force decreases until this position is reached. Hence after a sudden fixed strain, the force or stress decreases, this is called stress relaxation. The parallel combination of the spring and damper is the Voigt model. The strain cannot increase instantaneously after a step force is applied. The change in length of both elements is identical. The strain increases exponentially in time until the equilibrium position is reached ($\Delta l = F/k$). When the force is removed the strain cannot change instantaneously and will decrease exponential to the starting position. This lag in strain after a stress is applied is called creep. The Maxwell model can model stress relaxation only, the Voigt model only includes creep. Real materials exhibit both stress relaxation and creep and more complicated models are needed.
- 15) Resonance is the increase in amplitude of oscillation of a mechanical system exposed to a periodic force whose frequency is equal or very close to the natural undamped frequency of the system. (Source Wikipedia)
- 16) The damping ratio is one of the two parameters that determine the transfer function of the spring-mass-damper system. Its magnitude determines the degree of damping, energy dissipation and this determines the magnitude of the oscillation of the mass due to a periodic force. When this parameter is larger than one resonance is suppressed.
- 17) Undamped system: After a step displacement Δl the mass oscillates between position $l + \Delta l$ and $l - \Delta l$ with the natural resonance frequency ω_r . Elastic energy of the spring is converted into kinetic energy and vice versa. This oscillation of displacement of $2\Delta l$ lasts forever, there is no energy dissipation. When a harmonic force is applied at the resonance frequency, the amplitude increases with each cycle in principle to infinity. In real systems mechanical limits are reached, damage is likely.

16.2 Answers for questions in chapter 4

1. A parabolic flow profile is obtained for a tube with ideal Poiseuille characteristics. The assumptions are: laminar flow, Newtonian liquid, no-slip condition at the tube wall boundary, straight and very long rigid tube, static pressure gradient, mass less fluid, incompressible fluid, no gravitation effects.

2. For a pulsatile flow in a large diameter tube the inertance effect is more important than the resistance. The flow profile will resemble that of an inertance element which has a plug type flow profile.
3. A long fluid filled elastic tube can be modeled as a transmission line. Long tube length refers to the wavelength being shorter than the tube length. A transmission line can be modeled well with a series of discrete lumped elements, for instance simple L C R elements. The inertance is caused by the mass of the fluid. The compliance is related to the elasticity of the tube wall and the resistance can be related to viscosity of the fluid and visco elastic losses in the tube wall. To calculate the characteristic impedance, pulse wave velocity and damping parameters values of L,C and R per unit length need to be used.
4. Length =1m, radius=0.03 m, viscosity $\eta = 8.9 \cdot 10^{-4} \text{ Pa.s}$ at 25 °C , density (1000 kg/m³)
 Resistance: $R = \frac{8\eta l}{\pi r_0^4} \rightarrow R = \frac{8 \cdot 8.9 \cdot 10^{-4} \cdot 1}{3.14 \cdot 0.03^4} = 2798 \text{ Pa.s/m}^3$ (per meter)
 Inertance L: $L = \frac{\rho l}{A} = \frac{\rho l}{\pi r^2} = \frac{1000 \cdot 1}{3.14 \cdot 0.03^2} = 354000 \frac{\text{Pa.s}^2}{\text{m}^3}$ (per meter)
5. For air: viscosity $\eta = 1.82 \cdot 10^{-5} \text{ Pa.s}$ at 25 °C , density (1.2 kg/m³) this gives $R = \frac{8 \cdot 1.82 \cdot 10^{-5} \cdot 1}{3.14 \cdot 0.03^4} = 57.2 \text{ Pa.s/m}^3$ and $L=424 \frac{\text{Pa.s}^2}{\text{m}^3}$. The relative importance of the inertance is smaller for the air filled tube.
6. Double the radius. The resistance decreases with a factor 16, the inertance decreases with a factor 4. The compliance C is equal to $C = \frac{2\pi r^3 l}{Ed}$ increases with a factor 8. When the length is doubled all three double in magnitude.
7. The Young modulus E=5E5 N/m² and the wall thickness d is 0.001 m. This gives: $C = \frac{2\pi r^3 l}{Ed} = \frac{2 \cdot 3.14 \cdot 0.03^3 \cdot 1}{5.10^5 \cdot 0.001} = 3.37 \cdot 10^{-7} \text{ m}^3/\text{Pa}$.
8. The characteristic impedance of the tube is $Z_0 = \sqrt{\frac{L}{C}} = \sqrt{\frac{354000}{3.37 \cdot 10^{-7}}} = 1.03 \cdot 10^6 \frac{\text{Pa.s}}{\text{m}^3}$, the pulse wave velocity $v_p = 1/\sqrt{LC} = 1/\sqrt{354000 \cdot 3.37 \cdot 10^{-7}} = 2.89 \frac{\text{m}}{\text{s}}$. For an infinitely long tube a time dependent signal will propagate with the pulse wave velocity of 2.89 m/s, the ratio between pressure and flow at a certain position is equal to Z_0 . The resistance is more than 100x smaller than Z_0 , it was justified to neglect the resistive loss.
9. The hydrostatic pressure difference between head and feet is for a height difference of 1.7 m equal to $\Delta P = \rho g h = 1080 \frac{\text{kg}}{\text{m}^3} * 9.8 \frac{\text{m}}{\text{s}^2} * 1.7 \text{ m} = 18000 \text{ Pa} = 135 \text{ mmHg}$. When laying down the height difference is zero, there is no hydrostatic pressure difference between head and feet.
10. In a rigid tube the flow rate is determined by the pressure difference between the inlet and outlet. The flow through an elastic tube is more complicated, three pressures are of importance, the inlet and outlet pressures and the external pressure. The flow is determined by all three pressures via the two driving pressures and the transmural pressure. When there is flow through the tube the transmural pressure varies over the tube length. Hence the local tube diameter and local resistance vary over the tube length. The strong dependence of resistance on tube diameter causes a non-linear pressure distribution over the tube length. Three conditions can be recognized. In the first the inlet and outlet pressure are both larger than the external pressure, the tube does not collapse and the flow rate increases (although in a non-linear manner) with the pressure difference between in- and outlet. In the second case the inlet pressure is larger than the external pressure and the outlet pressure smaller. In this case collapse occurs in the tube near the outlet area and the flow rate does not increase with a decrease in the outlet pressure. This is the waterfall condition. Finally when the

external pressure is larger than both inlet and outlet pressure the tube is collapsed over the entire length and the resistance to flow is extremely high.

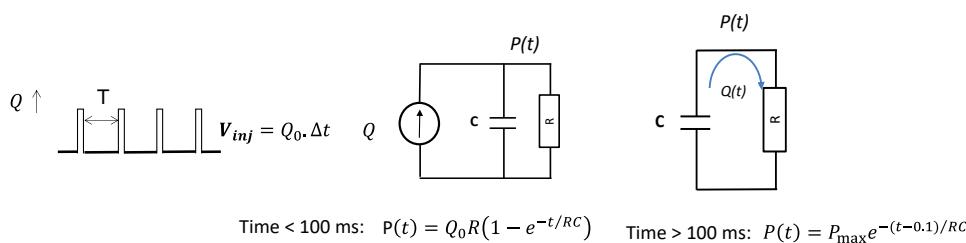
11. The transfer function H of a second order system is the equation that relates the amplitude and phase of the output parameter $y(t)$ for a given harmonic input signal $x(t)$ with angular frequency ω . The equation for amplitude transfer function in the frequency domain is:

$H(\omega) = \left| \frac{Y(\omega)}{X(\omega)} \right|$. In a second order mechanical system there is inertance L (related to mass), compliance C (related to elasticity) and resistance (energy dissipation due to viscous forces). The transfer function can be fully characterized by two parameters, the natural frequency $f_n = \frac{1}{2\pi\sqrt{LC}}$ the resonance frequency of the undamped system and the damping ratio $\zeta = \frac{R}{2\sqrt{LC}}$. When $\zeta < 1$ resonance can occur in the system and the resonance frequency is equal to $f_{res} = f_n\sqrt{1 - \zeta^2}$. For $\zeta < 0.7$ amplification can occur ($H(\omega) > 1$) in the frequency range around $f = f_n$ and for $f \gg f_n$ the amplitude transfer function goes to zero. At low input frequency ($f \ll f_n$) there is no amplification, the transfer function is approximately equal to one. The second order system behaves like a filter and has minimum distortion (i.e. frequency dependent amplification) occurs when the damping ratio $\zeta = 0.7$ and the resonance frequency is larger than the highest frequency in the bandwidth of the input signal.

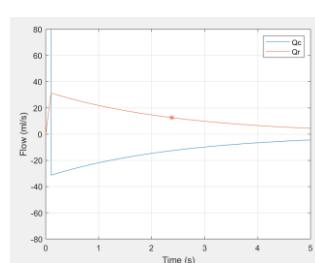
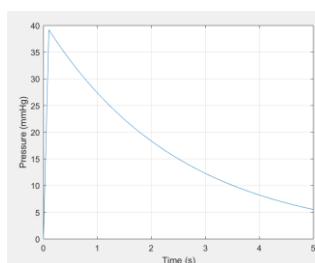
12. These networks are not equivalent in their input impedance. In the first and fourth network the input impedance is capacitive, this implies that the pressure at the input cannot change instantaneously, first flow followed by a pressure change at the input. For networks 2 and 3 the input impedance is inductive, this implies that the input flow cannot change instantaneously.

13. Use Ohms's law $R = \Delta P/Q \rightarrow R$ has the unit $mmHg \cdot s/l$. The compliance is $C = \frac{dV}{dP}$, C has the unit $l/mmHg$. For the inertance $\Delta P = L \frac{dQ}{dt} \rightarrow$ units of L are $mmHg \cdot s^2/l$.

14. Plots made using equations pages 73, 74 with given values of R, C , injection time, flow rate. Note that during injection there is also a flow through the resistor, the increase in pressure is therefore smaller than $\Delta P = \frac{80ml}{2 ml/mmHg} = 40 mmHg$. After $t=100$ ms the elastic energy stored in the compliance drives the flow through R .



R: 1.25 mmHg.s/ml
C: 2 ml/mmHg
RC = 2.5 sec
 $T_{inj} = 100$ ms



16.3 Answers to questions of Chapter 5

1. See Figure 5-1. A transducer converts one form of energy into another form, after conversion the output should have meaningful information for the user. A transducer consists of a sensing element (input signal), cables and connectors, analog and/or digital electronics (convert the input signal to the analog electrical domain, generates digital data) with and analog and/or digital output signal (waveform, numeric data).
2. The order of a system is the order of the differential equation that relates the input signal $x(t)$ to the output signal $y(t)$. For a zero order system there is a direct relation between the output and input signal, $a_0y(t) = b_0x(t)$ the proportionality constant $K = \frac{b_0}{a_0}$ is the gain. There is no time dependence, the response is immediate. A first order system is described by a first order differential equation $a_1 \frac{dy(t)}{dt} + a_0y(t) = b_0x(t)$, with DC gain equal to $K = \frac{b_0}{a_0}$, and time constant $\tau = \frac{a_1}{a_0}$. The maximum gain is the DC gain K , the gain decreases strongly for frequencies larger than the cut off frequency $f_c = 1/2\pi\tau$. A second order system is described by a second order differential equation $a_2 \frac{d^2y(t)}{dt^2} + a_1 \frac{dy(t)}{dt} + a_0y(t) = b_0x(t)$ with parameters DC gain $K = \frac{b_0}{a_0}$, resonance (angular) frequency $\omega_n = \sqrt{\frac{a_0}{a_2}}$ and dimensionless damping ratio $\xi = \frac{a_1}{2\sqrt{a_0 \cdot a_2}}$.
3. The damping ratio ξ should be larger than 0.7.
4. See text
5. The output of a clinical measurement is meaningful and actionable data in the form of a fully processed waveform (both analog and digital) and extracted numeric data. The waveform can be shown as is on the display of the monitor without further processing, the extracted numeric parameters can be used directly.
6. In a strain gage pressure sensor, four thin film resistors are formed on a thin flexible foil and are arranged in the form of a Wheatstone bridge. A pressure difference over the foil causes bending and results in a change of the resistance of the thin film resistors. Often material with Piezo resistance are used. For small bending of the foil the output signal of the Wheatstone bridge is proportional to the pressure difference. The output signal is amplified and after correction for the gain the output signal is an analog wave with amplitude and shape proportional to that of pressure wave.
7. A pneumotachograph is a very low Ohmic flow resistance which consists of many small diameter tubes in parallel combination. This suppresses turbulent flow in the device. Using Ohm's law the flow can be computed when the differential pressure over the device is measured. It can measure bi-directional flows. The dead space of the device cannot be neglected.
8. For both devices it is assumed that the fluid is incompressible and has no viscosity and gravitational effects are absent. Bernoulli's equation must be used for both sensors. Venturi: volumetric flow rate Q is the same in the wide (area A_1) and narrow tube (A_2). This gives $v_2 = v_1 A_1 / A_2$. Bernoulli gives $P_1 + \frac{1}{2} \rho v_1^2 = P_2 + \frac{1}{2} \rho v_2^2$, with density ρ . Combining these equations gives $v_1 = \sqrt{\frac{2(P_1 - P_2)}{\rho} \frac{1}{(A_1/A_2 - 1)}}$ and the flow rate $Q = A_1 \cdot v_1$. For the Pitot tube use again Bernoulli and realize that the flow velocity for the opening in the direction of flow is zero, the pressure there is the stagnant pressure P_1 . Assume that the flow velocity at the side port 2 is the same as the desired flow velocity v_2 . Bernoulli gives: $P_1 = P_2 + \frac{1}{2} \rho v_2^2$.

This gives for the fluid flow velocity: $v_2 = \sqrt{\frac{2(P_1 - P_2)}{\rho}}$. When a pitot sensor is used to measure the volumetric flow rate in a tube with diameter A it follows that $Q = A \cdot v_2$.

9. The height of the bobbin is determined by the balance between gravitational forces and forces due to the viscous flow along the bobbin in the conical shaped channel (via Ohm's Law and flow resistance R). The flow force depends on the viscosity, temperature and shape and dimensions of the conical tube. The scale depends the viscosity, sensor dimensions and temperature.
10. It is gas specific due to the dependence on viscosity and temperature.
11. See text
12. In continuous wave ultrasound the Doppler shift of reflected ultrasound waves is used to extract the flow velocity over the lumen diameter. This shifted frequency is equal to $f = f_0(\cos\alpha + \cos\beta) \cdot v/c$ and the flow velocity v can be extracted if the angles α, β are known. The Doppler shift can be measured by synchronous demodulation. This is a relatively easy-to-use technique for non-invasive determination of the flow velocity of blood at a specific site (i.e. carotid artery, radial artery). With some assumptions on the velocity profile it is possible to estimate the flow rate. When inertance is important plug flow can be assumed and estimation of the continuous volumetric flow is straight forward when the lumen diameter is known.
13. Pulsed ultrasound gives both imaging information and Doppler shifts at the imaged pixels in the blood vessel. The depth resolution should be much smaller than the lumen diameter. Hence the lumen diameter can be extracted from the imaging data and the flow velocity profile in the lumen can be determined. It is then possible to calculate the time dependent volumetric flow integrated over the lumen diameter.
14. When a high depth resolution is required a high ultra sound frequency is required. Attenuation of the beam increases strongly with frequency, hence the depth is limited first by the frequency used. Secondly the same transducer is used for transmission and detection. The transmitted bursts of US are separated by a period in which reflected signals can be sense. The time interval Δt between the burst limits the depth (i.e. $\Delta t \cdot c$) from which reflected signals can be detected.
15. The principle is similar to that used in phased array radar. The various elements in the array can transmit pulses with a predetermined phase difference. Due to interference between the different US beams only one direction remains. This direction depends on the phase difference. By varying the phase difference the angle of the ultrasound beam can be varied and can be scanned in the desired range.
16. Plethysmography is as sensing technique for measurement of (small) volume changes.
17. A cuff is a structure that consists of a thin outer layer that is flexible but relatively stiff and a bladder (fluid filled balloon) that is connected to the outer layer. The cuff can be wrapped around a limb (finger, arm, leg) and the cuff is inflated until it touches the limb. Volume changes in the limb lead to volume changes in the bladder, these bladder volume changes lead to changes in bladder pressure. When the fluid is air, and assuming that the outer layer is rigid, the pressure changes in the bladder can be converted into volume changes via Boyle's Law.
18. Photo plethysmography is an optical technique to study small changes in blood volume. Red and IR light is absorbed by blood. When there is pulsatile blood flow in a vessel its volume changes slightly and this is measured as changes in the intensity of transmitted or reflected light. For the operating principle see the text.
19. See text

20. The ppg signal consists of a DC and AC part. The DC part is about 100 times larger. The DC part depends on the non-pulsatile arterial and capillary blood volume, the venous blood volume, and tissue and bone absorption and transmission. During motion the LED emitter and photo diode change in position and the optical path of the transmitted/reflected will be different. Furthermore the non-pulsatile blood volume can vary leading to another large change in light absorption. The changes in signal due to motion can easily be 10 % of the DC signal and can be more than 10 times larger than the desired AC signal. Often the motion is complex and the motion artifact in the signal cannot be removed by filtering.
21. Changes in lung volume or changes in blood volume change the local resistance. These changes can be measured as variations in the bio-impedance.
22. Body currents are due to motion of ions and not electrons. The ion current needs to be converted to an electron current before it can be amplified in the ECG electronics. The ECG electrode is an electrochemical cell, most often an Ag/AgCl electrode. This electrode system has favorable properties for the ECG measurement. Changes in skin potential cause a reaction in this cell and electrons are released. The skin/electrode system can be modeled by a complex network of capacitors and resistors. At low frequency the capacitive impedance dominates the electrode impedance (~ 50 kOhm). At higher frequencies the capacitive part of the impedance becomes less important and contact resistances drops to low values (~ 100 Ohm).
23. The ECG signal is the change in time of electrical voltage difference between two electrodes at specific prescribed body locations. The body surface potentials are caused by currents flowing in the thoracic cavity that are related to the electrical activity of the heart. It can be measured as the voltage difference that is present between two ECG electrodes using a sensitive high-input impedance differential voltmeter with high gain (1000x) and high common mode voltage suppression (100 dB).
24. See text.
25. Besides the small ECG signal (pk-pk 1 mV), there is a DC electrode potential of 200mV-300 mV, 30-50kHz signals from bio-impedance, noise signals from coupling with 50Hz power lines and higher frequency EMI signals from for instance digital electronics and WiFi wireless networks. Furthermore pacing signals (pulse of a few ms width) from a pacemaker can be present. The 50Hz and digital noise can easily be larger than the ECG signal. For high fidelity measurements the required bandwidth is between 0.05Hz and 150 Hz. The DC signal can be removed by low pass filtering, digital and analog high frequency noise can be removed by high pass filtering. The pacing signal must be detected and removed from the signal. The 50Hz signal is in the bandwidth of interest, it can be removed by applying the so-called right-leg drive which aims to cancel the 50Hz component of the body and by a sharp notch filter at 50Hz.
26. The core temperature is the most important one, note that there is not a single core temperature, the most important ones are the brain temperature and temperature of the heart. This temperature is most relevant for clinical monitoring of the true body temperature as the skin temperature is much lower and depends a lot on environmental conditions.
27. Temperature sensing is done using thermistors, temperature dependent resistors, using a calibration table the value of the resistor can be used to obtain the temperature. Measurements of the skin temperature are easy. The core temperature requires either invasive measurements (catheters in bladder, large arteries, and esophagus) or minimally invasive (rectum). Alternatively a non-invasive zero-heat flux sensing method can be used. This sensor consists of a carefully designed and thermally insulated adhesive system with two temperature sensors, one at the skin, the other separated by a thin thermal isolator layer

- from the skin electrode. When the temperature difference is zero it is assumed that the measured temperature at the skin is equal to the core temperature.
- 28. Non-dispersive spectroscopy is a gas selective optical detection technique where there is no wavelength selection in the incident beam (this can be achieved for instance with a complex grating system but this is large and expensive). The desired wavelength is selected by the use of a filter before the photodetector. A system consists of a broadband light source (could be an incandescent lamp), an optical filter, an absorption cell with the gas mixture, a second optical filter and a photodetector (photo diode, thermal sensor). The spectrometer can also include a reference cell to correct for drifts in the light intensity and detectors.
 - 29. A capnograph is a sensor device that can measure the waveform of the partial CO₂ gas pressure as function of time during the breathing cycle. It should have a sufficient fast response rate and be sufficient sensitive. It can be made sensitive to CO₂ by measuring the absorption at a wavelength of 4.22 μm which is specific for CO₂ (a vibration band of CO₂).
 - 30. In main stream capnography the CO₂ partial pressure is measured directly at the airway opening. All expired air passes the CO₂ sensor. In side stream capnography a fraction expired air is pumped via small tube to the CO₂ detector. The advantage of mainstream capnography is that it measures the expired air directly and without delay, a drawback is that this measurement requires a secured airway as mask may be prone to leaks. The side stream device has the advantage that it can measure expired CO₂ also for a non-secured airway via an oral-nasal cannula. The disadvantage is the time delay caused by the pumping towards the CO₂ sensor (several seconds), pumping part of the inspiratory airway gas away from the patient, and condensation of water can block the small diameter tubes to the sensor. This requires special measures in material selection and water traps before the sensor.
 - 31. Filter-out disturbing wavelengths in the filter directly after the light source. Example CO and CO₂ absorption. Other options are to change the sensor gain factor when a known gas needs to be detected and analyzed. Finally multiple wavelengths IR spectrometers may be used to detect the different gases.
 - 32. See first part of section 10.2, note that this analysis is overly simplistic, there are more factors that influence the absorption of light, and also scattering is very important. Therefore empirical calibration methods are used to determine SpO₂.
 - 33. A pulse oximeter differs from an oximeter by measuring the pulsatile red and IR signals, in this way the saturation of arterial blood can be estimated. An oximeter would measure the total signal and would measure the average saturation of arterial and venous blood in the optical path.
 - 34. Methemoglobin has a large absorption in the red and IR bands, therefore when present in larger concentrations absorption from this molecule cannot be neglected and with two measurements at different wavelength three concentrations cannot be determined. Oximeters with more wavelengths are needed.
 - 35. The total concentration of hemoglobin and the saturation SO₂ need to be known, normally the dissolved O₂ can be neglected. This requires either a blood gas analysis or a pulse oximeter that measures at more than two wavelengths.
 - 36. The data in Figure 5-68 is needed. Low SaO₂ blood contains more deoxygenated Hb. Deoxygenated Hb has a large molar extinction coefficient at red light than HbO₂, the reverse is true at the IR wavelength. Hence blood with large SaO₂ absorbs a smaller amount of red light, the AC amplitude is smaller than for the low SaO₂ blood. Note that this is the true signal, smaller absorption corresponds with a smaller reduction red AC signal. The ppg signal is normally shown in an inverted manner. The IR AC signal at large SaO₂ is larger than at low

SaO₂ because the molar extinction coefficient of high SaO₂ blood is larger at the IR wavelength, hence more light is absorbed and the amplitude is larger.

37. At high SaO₂ the AC amplitude of the red light is small, the amplitude of the IR AC signal is relatively large. The ratio-of ratios is small. At low SaO₂ the amplitude of the red AC signal is large, the amplitude of the IR AC signal is reduced. The ratio R increases.

16.4 Answers to Questions of Chapter 6

1. In an obstructive lung disease the resistance of flow of air in the lung is increased, there is an obstruction of airflow into the lung. Often the airways collapse easily, increasing the resistance even further, this is most severe during active expiration. Exhalation of air is most problematic.
2. COPD has lung emphysema, accompanied by chronic bronchitis. Hence lung volume is very large, the lung is very compliant, and the area where gas exchange can occur is strongly reduced. The lung resistance is large, especially during active expiration. Airways collapse easily. In asthma the lung volume and elasticity are normal but the airway is in a state of continuous inflammation, mucus and smooth muscle contraction restrict the airway lumen diameter which increases the lung resistance. Exacerbations occur frequently, smooth muscle around the airways contracts even more and airway resistance increases even further.
3. In a restrictive lung disease the elasticity of the lung and/or chest wall is affected, the tissue is much stiffer, the compliance of the lung/chest wall combination is reduced. The airway resistance is normal.
4. See text
5. See Figure 6-24. Note the differences in total lung capacity, residual volume and airway resistance.
6. The lung spirometer – lung volume is constant, during breathing oxygen is consumed from the spirometer, expired CO₂ is absorbed. The spirometer volume decreases due to the O₂ consumption.
7. See text
8. The best method is whole body plethysmography because the measured lung volume also includes areas of the lung that are not ventilated (physiological dead space).
9. Spirometry, all lung volumes, airway resistance, forced inspiration and expiration tests, dead space when capnography is added.
10. In a whole body plethysmograph the airway resistance can be obtained when the airway opening pressure, lung pressure and air flow rate are known. The first and last term can be measured directly. The lung pressure can be determined during tests where a shutter is closed. The shift in lung volume when there is no flow can be recorded together with the reduction in lung pressure (is equal to AWP pressure when there is no flow). Using this data the lung pressure during quiet breathing can be measured.
11. Dead space is measured using either volume capnography (VCO₂) or the single breath washout method (SBW). In the latter method one breath of pure O₂ is inspired to maximum lung volume and the expired air composition and air flow rate are measured. Initially the expired air contains no CO₂ (VCO₂) or N₂ (SBW), this air comes from the dead space. When the partial pressure of CO₂ or N₂ reaches 50% of the final value the corresponding expired volume is dead space volume (or a similar method, more or the same). There are alternative methods such as the Bohr equation but this has not been treated in the course. The expired air from the alveolar region (P_{A-CO_2}) is diluted by the dead space air and the partial pressure in expired air is P_{e-CO_2} . To a good approximation the alveolar and arterial CO₂ partial pressures are equal, hence when the arterial CO₂ partial pressure (P_{a-CO_2}) is known from a blood gas test the dead space volume can be obtained $V_d/V_A = (P_{a-CO_2} - P_{e-CO_2})/P_{a-CO_2}$
12. See text
13. See text
14. This physiological term, the static compliance is measured at very slow respiration rate such that energy dissipated in the airway resistance and visco-elastic lung tissue can be neglected.

The dynamic compliance is measured when air flows at normal rates, in this case there is a P-V hysteresis loop and dynamic compliance can be inferred from this measured loop. (Note that this is a very sloppy definition and measurement). These parameters can also be estimated during PPV mechanical ventilation. The static compliance can be inferred from the tidal volume and difference between the plateau pressure and PEEP pressure.

15. See text, the basics is the flow in an elastic collapsible tube and the waterfall effect for certain pressure-volume conditions.
16. Via carbon monoxide diffusion, see text.
17. Consider the one lung model with a chest wall compliance and lung compliance. Given an elastic recoil pressure P_e , this pressure is divided over the lung and chest wall compliances. If the chest wall compliance is much smaller most of P_e will drop over C_{cw} and this will restrict the residual volume. A similar reasoning holds for the muscle pressure. After a maximum inspiratory effort the pressure drop over the lung will be small and the increase in lung volume will be smaller than the average case. Hence the maximum lung volume and residual volume will be decreases, the difference being the vital capacity.
18. Spirometry can be added in the ICU by monitoring pressure, air flow and volumes at the airway opening. Volume capnography is already available in many ventilation machines and in some monitors.
19. This is a two lung model with equal lung compliance C but two different airway resistances R_1, R_2 . Use complex impedances $Z_c = 1/j\omega C$ and use rules for addition of parallel impedances to derive an equation for the total impedance Z . Extract the imaginary part of the total impedance. You will find after a lengthy calculation that the imaginary part of the impedance is equal to: $Z_i = \frac{1}{\omega C} * \left(\frac{R_1^2 + R_2^2 + 2/(\omega C)^2}{(R_1 + R_2)^2 + (2/\omega C)^2} \right)$. You will find that when $R_1 = R_2$ the compliance C is equal to $Z_i = \frac{1}{\omega^2 C}$ as expected, with $C_{eq} = 2C$ and frequency independent, the equivalent capacitance is the sum of the two compliances. When $R_1 \neq R_2$ the equivalent compliance will be frequency dependent and is equal to: $C_{eq} = C * \frac{(R_1 + R_2)^2 + (2/\omega C)^2}{R_1^2 + R_2^2 + 2/(\omega C)^2}$ which is frequency dependent.
20. See text, these data give the most reliable and complete information on the oxygenation and composition of arterial and venous blood including oxygenation, oxygen content and concentration of CO₂ related species and pH of blood.
21. Bio-impedance, accelerometers, side stream capnography, flow sensors, microphone, cameras and radar are excluded because of the mobility of the patient. The respiband is omitted from the list because it is obtrusive and is used mostly for clinical studies as a reference device.

16.5 Answers to Questions of Chapter 7

1. The Windkessel model(s), see text for a description
2. There are many error sources. See page 38. One of the most important ones is that during critical care the arterial compliance and systemic resistance may change rapidly and that a new calibration is needed. The user may not be aware of this and large errors are made.
3. See text
4. See text
5. The recirculation part can be separated from the primary part by using an exponential extrapolation of the primary part where the recirculation part is negligible.
6. Invasive blood pressure is recorded in the aorta (near the aortic valve) or other large arteries and veins like the femoral, subclavian and radial artery and subclavian, vena cava, jugular vein. The blood pressure of the pulmonary circulation to the left part of the heart is not possible, the brain is very difficult.
7. Non-invasive measurements are done preferentially on the upper arm (brachial artery), other options are the wrist and fingers.
8. Wedge pressure is the pressure in the pulmonary artery when the flow is blocked by a small balloon. It is measured at the tip of a Schwann-Ganz catheter after inflation of the balloon. It measures the pressure at the output of the right ventricle which is a very important parameter for the function of the right heart and is important to diagnose pulmonary hypertension (a serious chronic disease).
9. The peripheral arterial pressure deviates from the central aortic pressure due to reflections of pressure and low waveforms in the transmission line like arterial tree. Furthermore the characteristic impedance is not constant in the arterial tree which causes further changes in pressure waves. Note that the systolic pressure is most affected.
10. A catheter-manometer system has second order filter characteristics due to presence of L,C and R in the system. The natural frequency is the resonance frequency of the system and the damping parameter is an important parameter that describes how oscillations decay after a step response. It is also an important parameter of the frequency response of a filter. For a flat response function up to 40Hz the natural frequency should be larger than 40Hz and the damping parameter should have a value of 0.7 as a trade-off between band width and distortion.
11. See text
12. Gas bubbles in blood can be very dangerous, CO₂ dissolves in blood and bubble formation is reduced.
13. See text, a too fast inflation or deflation increases errors as the change in cuff pressure is too large between two heart beats, a too slow inflation leads to pile up of blood in the venous system that may also cause pulsatile pressures in the veins and reduce blood flow distal to the cuff, i.e. makes pulse detection less reliable.
14. See text
15. This method relies on default values for the arterial compliance. This parameter changes when vasoconstriction causes blood flow centralization. Blood flow towards the peripheral smaller arteries is restricted, signals extracted from the finger, including ppg and pressure cannot be used anymore.
16. This method gives an estimate of the absolute values of the systolic and diastolic pressures, the Penaz method is well suited for the waveform. Normalizing the waveform to the measured NIBP removes the potential mean or DC pressure error.

17. This filter should be the inverse filter of that of the arterial system from sensor to finger. This holds for the magnitude. The phase of the various harmonics should also be corrected. Hence both amplitude and phase characteristics need to be taken into account.
18. See text, Note the calibration issue and changes of arterial compliance in time.
19. PTT is measure between two points in the arterial system, PAT is measured as a time difference between the r-peak of ECG and the foot of a ppg pleth signal.
20. One can measure PTT or PAT a specific moment in time, hence only one blood pressure value is available. Furthermore arterial properties change in time so using measurements at other times or during a treadmill test are prone to errors. There is also a large person to person variability, hence population averages cannot be used.
21. It is believed that velocity encoded MRI of flow in the central aorta is the most accurate method. The use of MRI restricts wide spread use. Measurements cannot be done at the bed side.

16.6 Answers to Questions Chapter 8

1. Device manufacturers have to comply with many medical standards for safety for both user and patient. Furthermore the device manufacturer optimize the system for the most demanding clinical use cases and workflow. In short simplicity and reliability in operation with guaranteed quality. The device should be trusted by the clinician, i.e. it must have guaranteed accuracy and reliability, low latency, warn when data integrity is in doubt, must have a simple user interface, must be easy to use, simplicity, rugged. Monitors are used for a long time therefore new systems should preferably have backwards compatibility. Use cases vary a lot, monitor requirements vary therefore and flexibility and the option to expand functionality of devices (including hot swap) and ability to add new modules is a key feature. Monitor requirements vary per department, hence a family of devices optimized in function and cost and use should be available. The manufacturer should also have offer solutions for central monitoring and data storage and connection to the hospital information system. The monitor must have internal storage for a specific time. Data should be linked to a patient. Context data and event data become increasingly important. Wired and wireless connectivity must be of the highest quality. Wired connectivity must offer an option for guaranteed quality of service. The device should have minimum down time for maintenance. Finally the device must be future proof.
2. Conflict between information and complexity. The information on the display should be presented in a systematic form that allows the user to directly obtain the most relevant information. There is often little time to gather the required information. Both waves and numeric data must be shown. Secondary contents should be easily assessable.
3. Very important. See text of the material of the introductory lessons.
4. Important principle, do not harm, risk-benefit analysis is needed. Monitoring is adapted to patient health condition. Two examples for surgery. For a relatively low risk surgery (e.g. young patient, knee injury during sport, local anesthesia) only the basic ASA requirements,, avoid invasive measurements (have risk but provide little benefit) as much as possible. For high risk surgery of patients (for instance trauma, cardiac, oncology ...) where hemodynamic status and stability can change rapidly, continuous beat to beat information of highest quality is mandatory. This requires invasive measurements (blood pressure, cardiac output, blood samples)
5. For a coronary care unit measurements are focused on the heart function and are preferably non-invasive. 3-5 lead ECG for continuous monitoring, 12 lead ECG for investigations and diagnosis, SpO₂ for blood oxygenation and pulse rate and automated oscillometric blood pressure. In the ICU high risk patients (unstable, fluid administration, medicines, blood samples, ...) with many different disease types are treated. Mechanical ventilation is common. Therefore most non-invasive and invasive measurements modalities are present. Minimally 3 lead ECG (waves and heart rate), chest bio impedance (respiration rate), SpO₂, temperature, oscillometric NIBP, invasive blood pressures (arterial, right heart, pulmonary artery) and continuous cardiac output are available. Mechanical ventilation is common, add airflow, airway pressure, capnography, gas composition. In the OR these measurements are also available but anesthesia specific monitoring modalities such a depth of anesthesia (EEG based) and specific anesthesia gases (N₂O, sevoflurane, ...) are added.
6. Normality and stability refer to health status and implies that the body systems function within normal physiological range. Hence the patient has normal values for the vital sign parameters, i.e. with values in the accepted physiological range(s) and variability that is typical for an average person (same gender, age, BMI, race, presence of chronic diseases...). When functions are abnormal organ injury and organ failure are possible. Therapy is

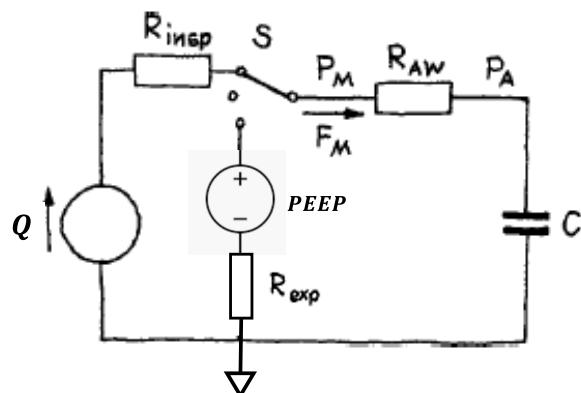
administered to prevent this from happening. Stability information can be obtained from a longer duration time series of the monitor parameters, this requires continuous measurements. Variations can occur slowly or sudden. For instance the health condition can change suddenly (e.g. a heart attack, cardiac arrest, shock).

7. The validity of waves and derived data presented to the user must be investigated before it is shown on the display. Signals are often noisy and have many artifacts and disturbances. The quality of data must be guaranteed, therefore analog and digital signal pre-processing is done to validate the signals and extracted data and features. This is preceded by signal validation, i.e. the device continuously monitors the quality of the signals (signal spectrum, signal to noise, artifacts) generates a so-called INOP alarm when signals deviate too much from normality. An example is the check of ECG electrode contact resistance (via bio-impedance) which can have a large impact on noise.
8. Use material from physiology, measurements and therapeutic devices material.
9. See text
10. See text
11. One measurement, continuous, 1 channel → $250 \text{ sps} * 2 \text{ bytes} = 0.5 \text{ kB/s}$. Take 2 channels per measurement (ECG, SpO₂, IBP,...) → 1 kB/s. 20 measurements → 20 kB/s. One day monitoring is 1.7 GB per day. The second case: 80 samples per second → $80 * 2 \text{ bytes} = 160 \text{ bytes per second}$. This corresponds to 14 MB per day. Third case sample rate is 30x lower than previous → $14\text{MB}/30 = 0.46 \text{ MB per day}$. This shows that storing waveforms requires very large storage capacity. Most data is stored at low rates mostly in the form of vital sign parameters.
12. Practical layout. Do not show more than is needed. Assume that screen has touch functionality. There must be a part that is required to enter a menu structure to change settings, measurements, frequency of intermittent measurements, alarm limits etc. this should be simple and intuitive. The remainder of the screen should show the waves with scales, each wave has its own color. Adjacent to the waves there should be space to display numeric values of the vital sign parameters that belong to the waves, use the same color for the data. The colors should be standardized for each parameter. On top could be a location that shows the type of alarm.
13. See text, figure
14. A clinical measurement includes the protection, analog pre-processing, digital to analog conversion, power and voltage generation, isolation from mains and data buses and subsequent digital signal processing and feature extraction. The wave output and parameter output can be used directly in the monitor. Hence it is a complete module that provides clinical useable data to the monitor. This is a complete functionality that is most often not present in physical measurements systems.
15. See text and figure
16. See text and figures
17. The main issue is the quality of service, for most common networks like WiFi, Bluetooth, GSM this cannot be guaranteed. Real time alarming cannot be guaranteed.
18. The functionality, safety and reliability requirements of a monitor cannot be guaranteed. The operating system is not real time, a monitor must have a real time operating system that is owned by the manufacturer. Furthermore the life cycle and support do not meet the requirements of a monitor device (10-20 year lifetime). It would also pose a severe burden on the software development team. Furthermore data privacy cannot be guaranteed.

19. Measurement isolation cannot be achieved, it is impossible to electrically isolate the different measurements for the kV demands and low leakage currents required at such voltages.
20. Monitors belong to specific groups, departments. Hence gaps in patient monitoring during transport / handover from department occur frequently, this is accompanied by loss of data. There is more demand for systems that can travel with the patient during the journey through the hospital. Wireless connectivity is still problematic, issues will increase due to the explosive growth in number of devices and continuous use. Technology gap with consumer devices increases. Cable clutter remains a very annoying problem, not easy to solve. Alarm fatigue remains major problem, reduction of false alarms a must. Data is underutilized. Smart alarms needed. Patients become more mobile (even in the ICU!), no good solutions for monitoring in the ward, home.

16.7 Answers to Questions of Chapter 9

1. The minute volume must be controlled, this is equal to the product of tidal volume and ventilation frequency. Lung protective ventilation to the patient is very important, this method corresponds with a limited tidal volume and reduced plateau pressure. Note that both airway resistance and lung compliance may change and ventilator settings may need to be changed regularly. For a pressure controlled device, the maximum airway pressure is controlled. The tidal volume depends on the airway resistance, lung compliance and inspiration time. The minute volume is not controlled directly, there is a risk that oxygenation of the patient is insufficient. Increasing the tidal volume requires either an increase in maximum pressure or an increase in ventilation rate. The former change increases the risk for lung injury, the second increases the risk for auto-PEEP (too short expiration time). For a flow controlled device the tidal volume can be controlled precisely but the maximum pressure is not controlled. For patients with high airway resistance, small lung compliance or with inhomogeneous lungs the maximum pressure may exceed critical limits and lung injury is possible. These problems can be resolved by monitoring of the airway pressure and flow.
2. A diagram of the model is shown below. It includes the mechanical flow source, the parasitic



resistances of the inspiratory and expiratory tubes, a switch to control inspiratory, hold and expiratory times, a pressure source in the expiratory tube for the PEEP pressure and the simplest one lung model of a patient. Note that in reality the tube resistances are not constant and flow and dependent (due to turbulence), the lung/chest wall compliance is also highly non-linear.

3. See text
4. See text
5. Clinicians estimate the airway resistance by inserting a short hold period between the inspiratory and expiratory phases of a breath in the flow controlled PPV mode. In the hold period the airflow is zero and AWO pressure equals lung pressure. The difference between airway opening pressure at the end of inspiration and in the hold period can be used to extract a resistance. Assuming that the pressure is measured at the airway opening the error due to the resistance of inspiratory and expiratory tubes is not included in the estimate. The measurement is not very accurate because the lung/chest wall is visco-elastic and the lung volume will increase during the hold period, this increases the reduction in pressure at the hold period. Furthermore the airway resistance is not constant, it depends on transpulmonary pressure and lung volume. At the end of an inspiration these resistances are smallest. Finally the flow resistance of the endo-tracheal tube is large and cannot be neglected.
6. See text, note that it is affected by airway resistance and visco-elastic losses.
7. See text

8. When the endotracheal tube is placed in the esophagus it is initially filled with CO₂ rich expired air. When a ventilation occurs the stomach is inflated with the fresh air and part of the CO₂ is removed, assume that the fraction is equal to α . Each breath this same fraction is removed from the stomach/tube combination. Hence the CO₂ pressure as function of the number of breaths n is equal to $P_{CO_2}(n) = P_{CO_2}(0) * (1 - \alpha)^n$. This is an exponential decrease in pressure. This illustrates why it is important to use capnography during mechanical ventilation as insertion of the tube in the esophagus is lethal when it is not noticed (and the error is easily made).
9. The excretion of CO₂ per minute is VCO_2 , it is typically 200 ml per minute. When the ventilator produces a lung minute volume of 5 L/min the average CO₂ partial pressure during expiration is $P_{CO_2} = (200 \text{ ml}/4000 \text{ ml}) * 760 \text{ mmHg} \approx 39 \text{ mmHg}$. Assuming that VCO_2 is not changed the partial CO₂ pressure in expired air decreases when the minute volume is increased (more dilution) and increases when the minute volume is reduced. Note that VCO_2 changes with ventilation minute volume and eventually the partial pressures of CO₂ in blood will change but this takes time.
10. As shown in question 10 the ETCO₂ value depends on minute volume and is only a good representation of the arterial CO₂ partial pressure when the minute volume is just right. Note that when the ventilation is not optimal VCO_2 will also change, this can lead to hypo- or hypercapnia in blood.
11. The minute volume has to be increased, as shown above this will reduce ETCO₂ and increase the diffusion of CO₂ from blood to the lung, increase VCO_2 and reduce the partial pressure in blood. This increase in minute volume can be achieved by an increase in tidal volume and/or ventilation rate. The tidal volume increase can be achieved by an increased flow setting or increase of the maximum pressure.
12. Lung protective ventilation aims at reducing ventilator induced lung injury, especially for patients with severe lung diseases such as ARDS, COPD and pneumonia. To reduce barotrauma the plateau pressure must be limited to 30 cmH₂O, to reduce volutrauma the tidal volume should be adjusted to the patient lung volume (6ml-8ml per kg PBW). Since a certain minute ventilation volume has to be achieved for oxygenation and CO₂ excretion a small tidal volume and high ventilation rate are to be preferred.
13. Assisted ventilation is used when there is a spontaneous respiratory effort of the patient but he/she is still too weak to breathe independently. The advantage is that the ventilation rate can now be controlled by the patient reducing the probability of over- or under-ventilation. The ventilator detects an inspiratory effort of the patient by a small reduction in the airway pressure or increase in the airway flow and subsequently a trigger of the machine is generated and an assisted breath is delivered. There are several variations of techniques possible.
14. Pressure support ventilation (PSV) is one of the most used assisted ventilation techniques. In PSV the patient triggers all breaths, this is followed by a trigger of a PPV breath with a preset maximum pressure and often cycled off at a preset value of the maximum airway flow. In this way the patient can control the breathing rate and minute volume.
15. Asynchrony between the patient and the ventilator occurs when there is a mismatch between the patient and ventilator in terms of breath delivery timing.
16. Measurements of the pressure, flow and volume at the airway opening are needed but may not be sufficient as especially detection of the beginning of a spontaneous expiration effort is difficult to detect. For more certainty an esophageal pressure sensor can be inserted, this sensor pressure is proportional to intrapleural pressure which shows the patient effort. Other sensors measure muscle activity (EMG) or neuronal activity.

17. This is a normal assisted breath as both the pressure drop in the airway and increase of flow are visible before the ventilator is triggered.
18. The airflow causes a pressure drop over the inspiratory and expiratory tubes. From Ohm's Law this is equal to: $\Delta P = Q * R_{i,e}$, furthermore airflow is turbulent and the tube resistance is given by $R = R_0 + k \cdot Q$. Looking to the flow waveform it appears that the flow is initially peaked, hence both the flow and flow resistance are largest and the pressure drop is large, the AWO pressure is much smaller than the pressure in the ventilator. At the end of an inspiration the flow is small and the pressure drop is also much smaller, the pressure difference between the ventilator and AWO is reduced. During expiration airflows out of the lungs and lung pressure is larger than PEEP pressure, pressure at the AWO is larger than PEEP.
19. PEEP is needed to recruit collapsed lung volume, prevent collapse in the lung, opens the airway and to improve oxygenation. Furthermore these patients suffer from a high airway resistance which is smaller at large lung volumes.
20. See text
21. To ensure that the endo-tracheal tube is inserted in the trachea capnography is required (correct capnogram should be observed, exponential decline of PCO₂ is observed when tube is inserted in esophagus). Leaks or disconnection detection require flow and pressure measurements at the airway opening. Leakage of the ET-tube cuff can be observed by measuring differences between inspiratory and expiratory volume when the cuff is inflated and deflated. When the ET-tube cuff functions correctly differences in the volume waveform between inspired and expired volume are an indication of leaks in the system. When tubes are disconnected there is no flow and pressures are atmospheric.
22. Closing the loop is jargon for a fully automated ventilation in which all ventilator settings are set and controlled by the machine and human intervention is minimized. It originates from the term control loop in a servo system.
23. See text

