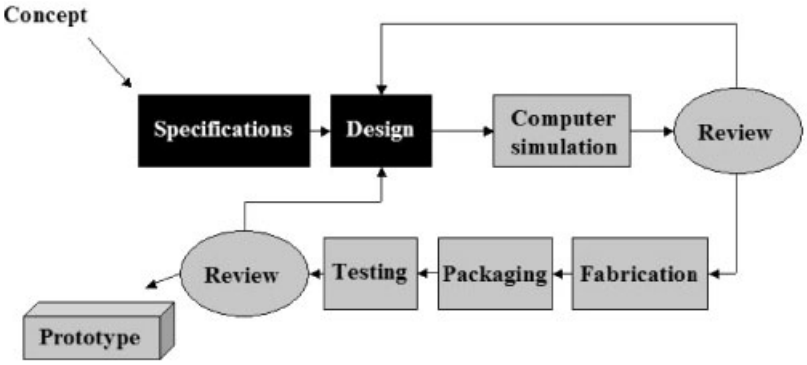
4
Microfluidics – Components

Jörg P. Kutter, Klaus Bo Mogensen, Henning Klank, and Oliver Geschke



4.1
Valves and Pumps

Among the most important fluid-handling elements are pumps and valves. Pumps are devices to set fluids into motion, and valves are designed to control this motion and, for example, define a preferred direction for the motion of the fluid. The engineering community has developed a wealth of pumps and valves based on a large variety of principles. Many of these principles and implementations can also be used in the microdomain, but additionally new designs have appeared making clever use of the different behavior of liquids and mechanics on small scales. It is far beyond the scope of this book to cover all proposed and tested micropump and microvalve designs. In fact, new designs appear almost daily. A good overview of many pump and valve principles adapted or developed for microfluidic devices can be found in, e.g., [1–4]. In this section we look at but a few selected examples of how pumping and valving are implemented in microsystems. A special section (4.4) is dedicated to pumping by electroosmosis, because this is a unique and widely applied principle deserving special attention. The remainder of this section deals with pumps generating so-called hydrodynamic flow, typically based on a pressure difference between the inlet and outlet of a channel.

Microsystem Engineering of Lab-on-a-chip Devices
O. Geschke, H. Klank, P. Telleman

ISBN: 3-527-30733-8

A valve is a device within a fluidic system in which flow is allowed in one direction but suppressed in the opposite direction, thus introducing directionality into the flow. Valves are often classified by whether they work by themselves (i.e., by utilizing energy from the flow) or whether they need external energy to function. The former are called passive or check valves, whereas the latter are called active valves. The wish-list of ideal characteristics that a valve should have includes zero leakage (when it's closed it's closed, when it's open it's open), zero power consumption (obviously not true for active valves), zero dead volume (should not introduce extra volumes, which negatively affect the performance), infinite differential pressure capability (a little bit of extra pressure from one side should open it, a little bit of extra pressure from the other side should close it), zero response time (no delays), insensitivity to particulate contamination, ability to operate with any fluid, and so on. Unfortunately, no single valve meets all these requirements, in other words, all valves have some flaws, which need to be taken into account when operating them.

An example of a passive valve is the cantilever type (Fig. 4.1), here, a thin strip of silicon that bends when enough pressure is applied from one side. From the other direction bending is restricted by the valve seat. Depending on the exact design and operating conditions, these kinds of valves are more or less leaky. However, they are relatively simple to make, do not require external energy for operation, and have a fairly fast response time. There are also passive valves, which do not rely on mechanical action, but rather make use of forces such as surface tension. These types of valves are typically one-time-use or burst valves, because their functioning depends on an air–liquid interface, which is typically present only when first filling a microchip. A simple implementation of such a valve is a restriction in the channel. As the air-liquid interface approaches the restriction the

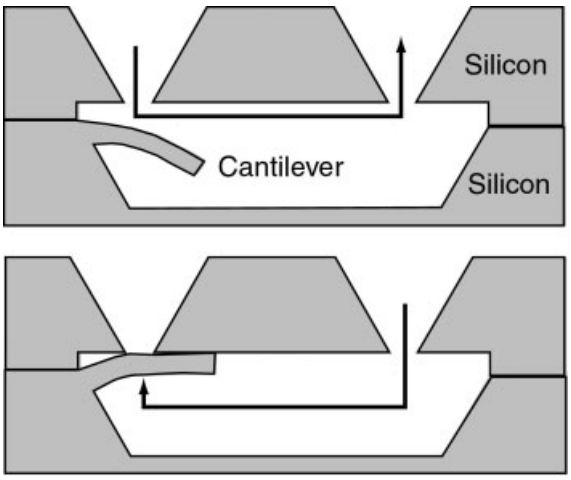


Fig. 4.1 A check valve using a cantilever as a valve flap (adapted from [1]).

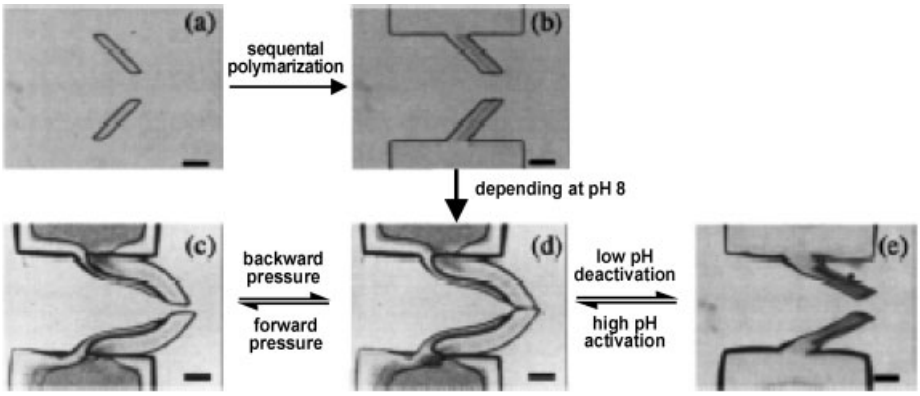


Fig. 4.2 Example of a valve fabricated of pH-sensitive hydrogels, which have different swelling behaviors depending on the pH of the solution (reprinted with permission from [5]. Copyright (2001) American Institute of Physics).

liquid cannot penetrate into the next channel segment because of the increased surface tension around the restriction. Only when a higher pressure is applied will the liquid break or burst through the restriction into the next section of the channel network. Similarly, defined patches on the channel bottom, which are chemically treated to have a hydrophobic surface can prevent aqueous liquids from flowing until sufficient pressure is applied to overcome the surface tension. Recently, another type of (basically passive) valve has appeared, based on hydrogels. Hydrogels are materials that can change some physical parameter (typically, their volume) based on some physical or chemical stimulus. One example uses small strips inside a channel made of different hydrogels, one of which is pH sensitive (Fig. 4.2) [5, 6]. Depending on the pH of the solution flowing through the channel, the hydrogel swells or stays in the shrunken state. If it swells the flaps block the passage of solution in a certain direction. Because the valve in a way 'senses' the solution and acts accordingly, these valves have also been called intelligent or adaptive valves.

Active valves require an actuator to provide a mechanical action, which moves a part of the valve to close or open the flow passage. Valves can be designed to be 'normally open' or 'normally closed', meaning that this is the state they are in without actuation. For reaching and maintaining the opposite state, energy needs to be applied. A few designs have two stable states (open and closed) and need energy only for transition between the two states (so-called bistable valves). The actuation principles used are various, including pneumatic (using compressed air), thermopneumatic (using heated fluids), piezoelectric (using special materials that expand when an electrical voltage is applied), electrostatic (using electrical attraction or repulsion), shape memory alloy (metals that change shape under temperature changes but 'remember' and revert to their original shape), and electromagnetic actuation, to name but a few. Actuators are typically characterized by four

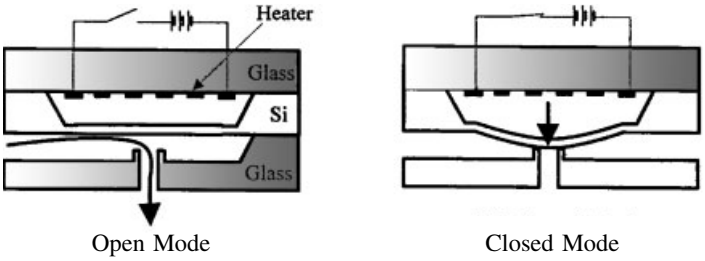


Fig. 4.3 An active valve using a heating chamber with a working fluid and a diaphragm (reprinted with permission from [3]. Copyright (1998) Springer Verlag).

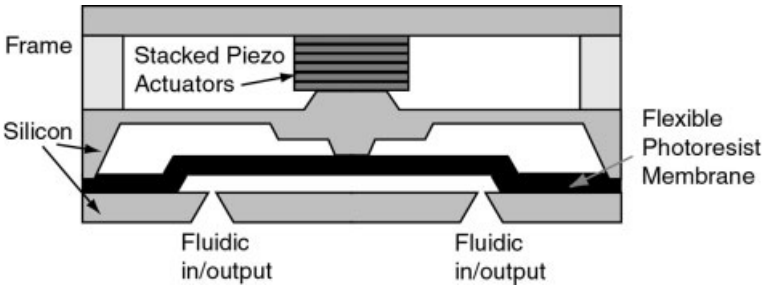


Fig. 4.4 An active valve having a piezoelectric stack as the actuation element (adapted from [1]).

criteria: the pressure they can build up, the stroke displacement they generate, their response time, and their reliability. Again, some actuators perform better in one category but less well in another. For example, piezoelectric stacks can produce large pressures, but not very impressive strokes. Electromagnetic actuators, on the other hand, can give large displacements but only relatively small pressures. A valve with a thermopneumatic actuator is depicted in Fig. 4.3.

A range of heating elements can heat a working fluid in a chamber on top of a diaphragm (50 μm thick silicon). The expansion of the fluid bends the diaphragm, which presses onto a valve seat sealing off the liquid flow. A different design with a piezoelectric stack as the actuator is shown in Fig. 4.4. Here, a thin, flexible photoresist membrane is included in the design to provide a tighter seal, i.e., less leakage. Softer gasket materials are supposed to be better suited to handling particulate contamination, which can prevent proper functioning of a valve especially when only hard materials are used.

Many different valves are available and imaginable, all using different designs, materials, and actuation principles. There is no optimum valve, but instead valves have to be chosen carefully depending on the intended use of the microsystem. Questions such as: what kind of solution is being moved through the system, i.e., comes in contact with the valves?, how often do we need the valves (many times per minute, only a few times in total)?, can we afford to spend a lot of energy on

actuating the valves? and so on have to be considered before designing or choosing a valve. Often, microfabricated valves still have too many flaws or drawbacks and researchers revert to off-chip, conventional miniaturized valves. The same can apply for pumps.

In its most common version, a microfabricated pump consists of a pump chamber and two valves: an inlet valve and an outlet valve. The pump chamber is mostly a combination of a diaphragm/membrane and an actuator to displace the membrane or diaphragm. This displacement, together with the action of the valves, results in alternating flux into and out of the chamber, i.e., pumping. The valves used can be any of the aforementioned valves, passive or active, as long as they can be used repeatedly over a long period of time. There are also valveless pumps as discussed below.

The requirements for a microfabricated pump can be as manifold as its intended uses. A typical list of requirements for a microfabricated pump to be used in a microsystem for analyzing wastewater contained the following points [7]:

- as little pulsation as possible
- flow rate controllable over a certain range (e.g., 0.25–10 $\mu L\ min^{-1}$)
- leak rates under 1 $nL\ min^{-1}$ at counterpressures of 1–10 kPa
- sensitivity to counterpressures of 1–10% kPa^{-1}
- flow rate precision over several days within 1%–3%
- resistance to aggressive chemicals over a longer period of time at elevated temperatures
- production costs as low as possible

Such requirements are made to assure accurate analytical results, reliable operation, and economic viability of a microsystem utilizing such a pump. At the same time, many of these requirements are extremely hard to meet, and meeting some of them often means seriously compromising other requirements. For instance, to be able to withstand aggressive chemicals (pH 11–12) over a long period at higher temperatures (about 40–50 °C) almost prohibits the use of silicon as a material for the microfabricated pump. Alternative materials need to be investigated, for which only a limited range of machining possibilities exist, and which have different mechanical properties (see chapter 7 and chapter 8). Such materials might also prevent the application of already developed and tested silicon-based actuators.

Alternatively, protective coatings could be applied, making additional processing steps necessary and also affecting the mechanical performance of the device.

As mentioned above, microfabricated pumps employ the same type of actuators as active valves. The combination of actuators chosen for the pumping chamber and the valves, respectively, determines the performance of the overall pump, at least with respect to generated pressure, stroke displacement, response time, and reliability. Additionally, the achievable flow rate and the amount of pulsation are criteria for microfabricated pumps. In a pump with piezoelectric actuation of the pump chamber and two ring-type passive valves (Fig. 4.5), even though piezocrystals can be driven with a very high frequency, the employed type of passive valves do not respond extremely fast and therefore limit the range of possible driving fre-

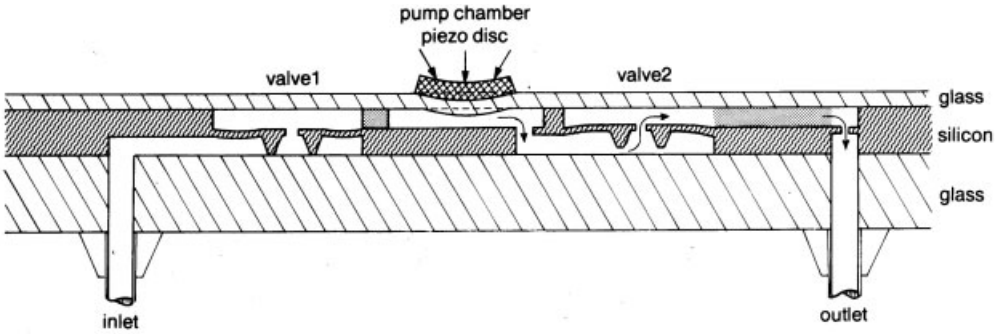


Fig. 4.5 A relatively simple pump design involving a pump chamber with a diaphragm, a piezoelectric actuator, and two ring-type passive valves [4] (courtesy H. van Lintel, EPFL).

quencies. Cantilever-type passive valves have a faster response, but provide a less tight seal and are more prone to leaking. A pump design that works entirely without valves, thereby reducing the amount of moving, and potentially failing, parts is shown in Fig. 4.6. Here, a pump chamber with a membrane is actuated as in many other pump designs. However, instead of inlet and outlet valves, asymmetric restrictions separate the pump chamber from the rest of the channel system. Depending on the exact geometry of these asymmetric restrictions, fluidic resistance in the two possible directions differs, thereby rectifying the flow into a preferred direction. The asymmetric restrictions function as either a diffuser or a nozzle, depending on whether flow is into or out of the chamber, which is why they are also referred to as diffuser–nozzle elements. Obviously, a micropump with diffuser–nozzle elements always has a relatively high degree of leakage, but is potentially more reliable because it has fewer mechanical parts.

All of the many designs of microfabricated pumps with actuators suffer from delivering pulsating flow, which is unavoidable because all actuators use a reciprocating movement. An elaborate design in which several of these pumps work together out of phase could improve the situation somewhat.

However, other (simpler) approaches to micropumps can deliver a steady, nonpulsating flow, much like a conventional syringe pump. One possibility is to use a constant-pressure gas reservoir, which presses on the liquid and displaces it. The gas reservoir could be pressurized by a microfabricated (pulsating) gas pump and regulated by valves and pressure sensors. One of the additional advantages of such a design is that the potentially corrosive liquid never comes in contact with the micropump itself. Pressure differences between two ends of a channel can also be generated by simply creating a difference in the liquid filling level or by placing a small and a large droplet at the ends and utilizing the larger surface tension of the smaller droplet. If channels need to be filled only once (also called priming) capillary action can be exploited to move liquids into channel networks. Even though the above-mentioned approaches are relatively simple, more advanced designs are required if flows of constant magnitude are desired.

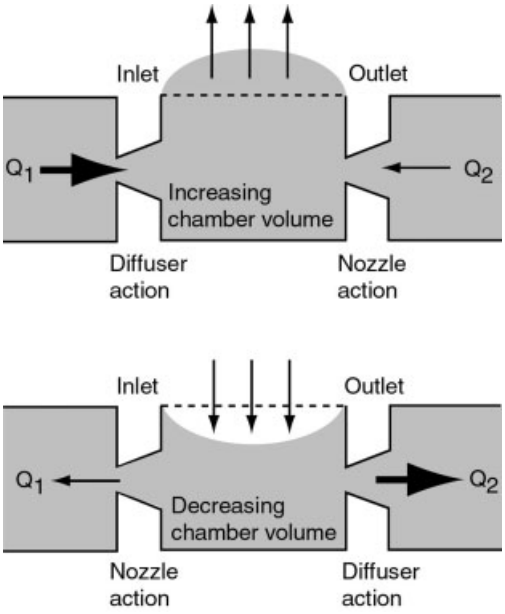


Fig. 4.6 Diffuser–nozzle designs to direct flow. Any of the mentioned actuation principles can be used to actuate the pump (adapted from [1]).

Another very interesting way to provide nonpulsating flow is to use centrifugal force. Imagine microchannels fabricated onto a CD-like disc and arranged in a mostly radial pattern (Fig. 4.7). If the disc is set spinning, centrifugal force moves the liquid inside the channels from the center of the disc towards its rim. Valving in this case is achieved with, e.g., burst valves or hydrophobic patches. At initially low angular speeds the liquids do not have the necessary energy to break through these valves, only as the disc's rotational speed is increased do they burst through. Rather elaborate liquid handling has been demonstrated on such centrifugal-force-driven systems [8–10].

Finally, methods utilizing the phenomenon of electroosmosis yield nonpulsating steady flows and typically do not require any moving parts, making them very interesting for a number of reasons. This is explained in more detail in the next section.

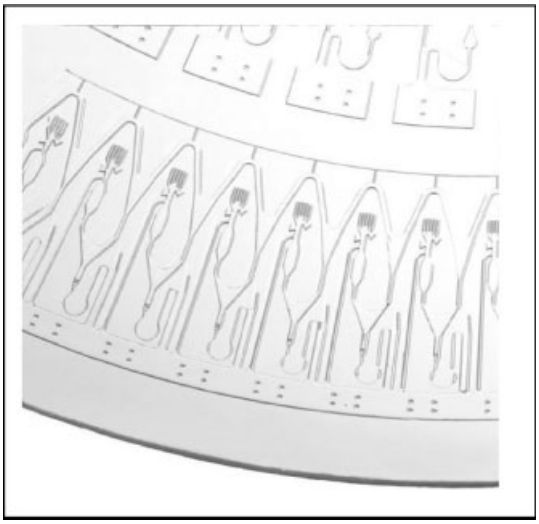


Fig. 4.7 Detail of a CD containing fluidic structures; hydrophobic patches function as valves (courtesy Gyros AB, Sweden).

4.1.1 Moving Liquids by Electroosmosis

At every interface, material characteristics such as charges and forces are not balanced in the way they are inside the bulk, giving rise to several phenomena, including electroosmosis. Let us take a closer look at the interface between a glass capillary or channel and a buffer solution. At the surface of the glass are silanol groups (–Si–OH), which, depending on the pH of the buffer solution, are deprotonated to a greater or lesser extent. Deprotonation results in charge separation, with the negative charges (–Si–O^-) immobilized on the wall and the protons immediately adjacent to the wall. Most of these positive charge carriers are also immobilized, due to strong electrostatic interactions with the negatively charged wall. However, another, diffuse layer of, again, negative and positive charge carriers forms further into the bulk of the solution. This arrangement of charge carriers, also denoted an electrical double layer, gives raise to an electrical potential between the wall and the bulk of the buffer solution. It is important to remember though, that the entire system is still electrically neutral. A schematic depiction of the double layer model and the resulting potential as a function of the distance from the wall are shown in Fig. 4.8. The potential at the shear plane between the fixed Stern layer and the diffuse Gouy–Chapman layer is called the zeta potential (ζ potential) [11–13].

The zeta potential is strongly dependent on the chemistry of the two-phase system, i.e., the chemical composition of the wall (material, dynamic or permanent coating, etc.) and the chemical composition of the solution (pH, ionic strength, additives, etc.), as well as on the temperature [13]. If we now apply an external voltage across the buffer solution we can induce a bulk flow of the entire buffer. As an elec-

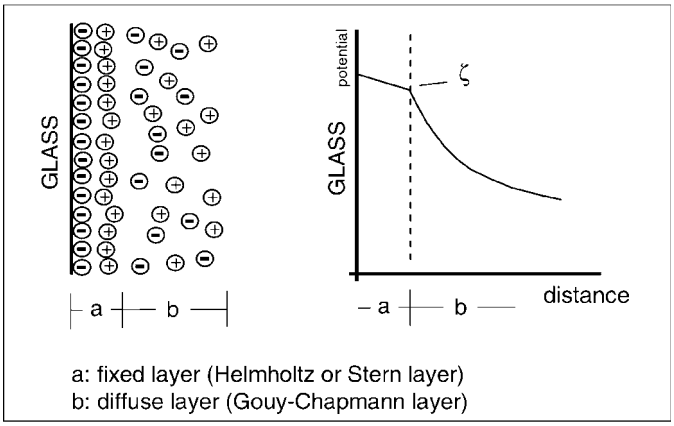


Fig. 4.8 Schematic of the electrical double-layer model and the electrical potential associated with it.

tric field is applied, the two types of charge carriers are dragged by electrophoretic forces towards the respective electrodes. However, since the negative charges from the deprotonated silanol groups are immobilized at the wall there is a surplus of mobile positive charges in the solution (while the entire system is always electrically neutral). These charges (mainly hydrated protons) drag the entire liquid column towards the cathode. The resulting flow is called electroosmotic (or sometimes electroendoosmotic) flow. We can define an electroosmotic mobility, μ_{eo}, as follows:

$$\mu_{eo} = \frac{\zeta_0 \varepsilon}{4\pi\eta} \tag{4.1}$$

with ζ_0: zeta potential; ε: dielectric constant; and η: viscosity. Once an electric field, E, is applied, this results in the following electroosmotic velocity, v_{eo}:

$$v_{eo} = \mu_{eo} E \tag{4.2}$$

Typical field strengths used lie in the hundreds of volts per centimeter, resulting in linear flow velocities on the order of 0.1 to a few mm s^{-1}.

To summarize at this point: it is obviously possible to create a flow (induce pumping) by filling a microchannel with a buffer solution and applying a suitable voltage at the channel ends. This relatively easy way is clearly an advantage of electroosmotic pumping, especially compared to the effort it takes to fabricate and run some of the microfabricated pumps presented in section 4.1. Another great advantage is the resulting flow profile. Contrary to hydrodynamic flows, where one finds a parabolic distribution of the flow velocities with the largest velocity at the center of the channel and zero velocity at the walls, electroosmotic flow is generated close to the wall and therefore produces a plug-like profile with a very uniform velocity distribution across the entire cross section of the channel. The proxi-

mity to the wall can be assessed by calculating the thickness of the double layer, the so-called Debye length, λ_D:

$$\lambda_D = \sqrt{\frac{\varepsilon RT}{2F^2 z^2 c}} \tag{4.3}$$

where R: universal gas constant; T: absolute temperature; F: Faraday constant; z: charge number; and c: concentration. For a standard borate buffer ($\varepsilon = 78.3\ \varepsilon_0$; $c = 100$ mM, $z = 1$) at standard conditions, it is straightforward to show that λ_D is on the order of 1 nm, i.e., more than a thousand times smaller than the channel width or depth. Therefore, virtually the entire cross section has a uniform velocity. Experiments have revealed parabolic and flat flow profiles by imaging a narrow band of released caged dye within a capillary under hydrodynamic and electroosmotic flow conditions, respectively (Fig. 4.9) [14].

The main disadvantage of electroosmotic flow is its strong dependence on the chemistry of the system, which is of course a consequence of the strong dependency of the zeta potential on the chemical state of the system. This dependency makes electroosomotic flow hard to control: every change in pH, dielectric constant, concentration, etc. due to reactions or mixing processes has an immediate effect on the magnitude of the electroosmotic flow. Using buffers can alleviate this situation to some extent and allows for reproducible, constant flow velocities, but only within certain limits. On the other hand, this strong dependency also means that there are many parameters that can be exploited to control and manipulate electroosmotic flow. For instance, lowering the pH reduces the magnitude of the electroosmotic flow and can even suppress it totally. The same can be achieved with a range of chemicals, which can be used to dynamically or permanently coat the channels walls and thereby alter their chemical behavior [15–17]. Here, it is even possible to reverse electroosmotic flow and induce bulk flow towards the anode. Suppression and reversal of the electroosmotic flow are valuable handles for tweaking the performance of electrophoretic separations (see also section 10.4). Finally, electroosmotic flow can be influenced by an external electric field applied through very thin insulating channel walls [18].

Apart from the direct pumping by electroosmosis, this phenomenon can also be used in an indirect way [19–23]. Only two examples of such implementations are briefly described here. Let us look at a T-type channel intersection (Fig. 4.10). If we apply a voltage from reservoir 1 to reservoir 2 we find that all the liquid follows this path, i.e., the fluid flow follows the current flow. If we now chemically modify the surface in the side-arm labeled C2, so that electroosmotic flow in this section is suppressed, but still have the current flowing as before, the following happens: at the intersection the fluidic flow generated up to this point wants to follow the current flow into the side-arm. However, this is now almost impossible because electroosmotic flow is chemically suppressed in this part. Consequently, pressure builds up at the intersection and the fluid follows the path of least resistance, i.e., becomes divided into a part flowing towards C2 and a part flowing into the field-free region towards C3, the ratio depending on the ratio of the actual

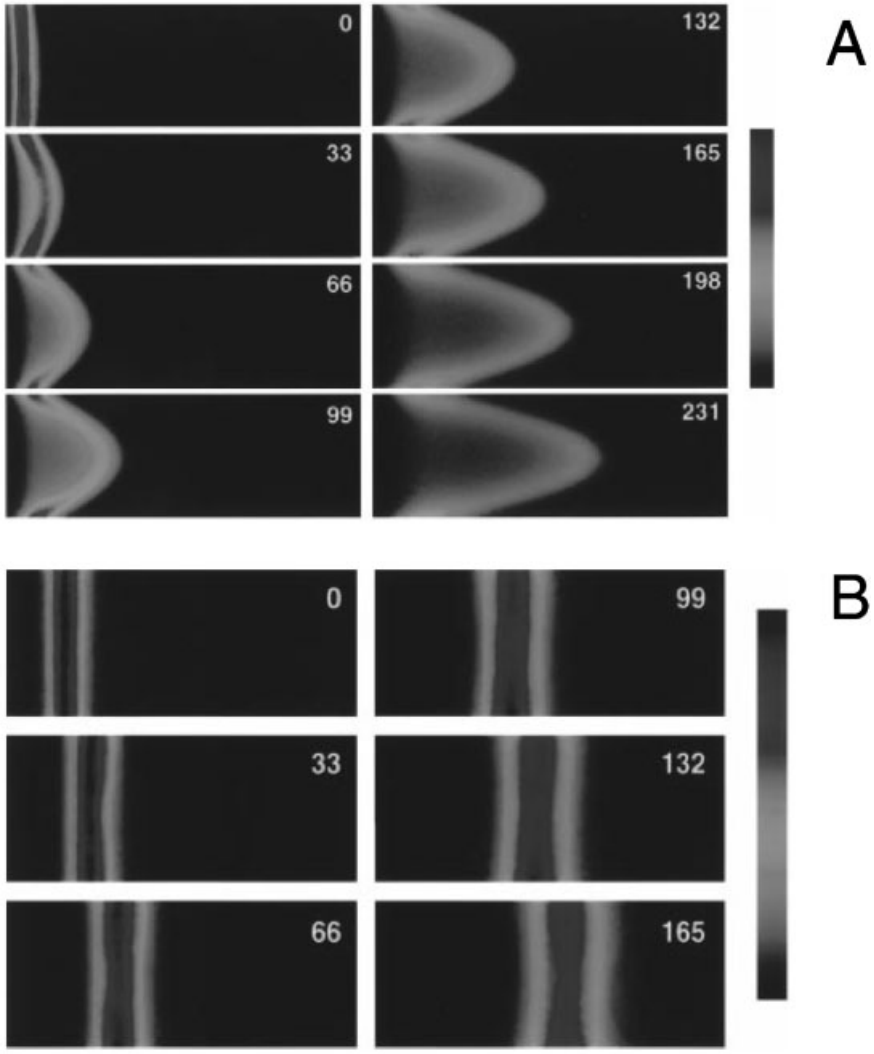


Fig. 4.9 Development of flow profiles imaged by using caged and subsequently released fluorescent dyes. A) hydrodynamic flow; B) electroosmotic flow (reprinted with permission from [14]. Copyright (1998) American Chemical Society).

flow resistances in these two parts. In this fashion, we have induced hydrodynamic pumping (with a parabolic flow profile) by means of electroosmotic flow. We have thus utilized the relative simplicity of generating electroosmotic flow to facilitate liquid movement in a different, field-free part of the channel network. In many applications it is preferable to avoid exposure to high electric fields, e.g., when working with living cells. Electrodes used for applying the voltage do not necessarily need to be placed at the ends of the channels, but can be placed at al-

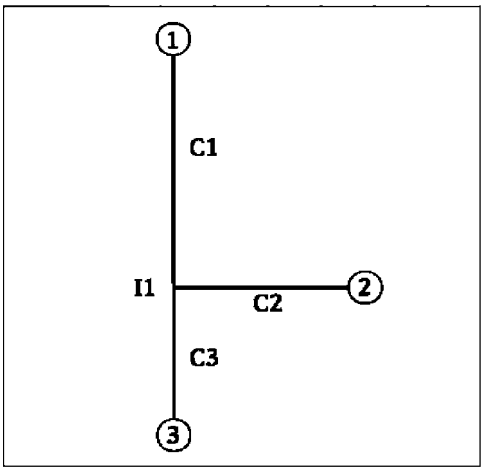


Fig. 4.10 Channel layout for electroosmotically-induced hydrodynamic pumping. C2 is chemically surface-treated to suppress electroosmotic flow (reprinted with permission from [19]. Copyright (1998) American Chemical Society).

most arbitrarily short distances from each other somewhere inside the channels. In these arrangements, the electric field is applied only between the electrodes, and the rest of the channel is field-free. Electroosmotic flow is also generated only in the space between the electrodes, but, because the liquid has to move somewhere, this induces hydrodynamic flow outside the region between the two electrodes [22]. A variant of this setup uses porous packings or porous polymer immobilized between two electrodes within the channel to increase the surface area and thereby the magnitude of the electroosmotic flow [24–26]. However, whenever electrodes are placed inside channels, care has to be taken to allow gas bubbles, which might be generated by electrolysis at the electrodes, to leave the system, because they can grow and interrupt current and fluidic flow.

4.1.2 Mixers

The importance of mixing in chemical microsystems is discussed in chapter 3.3. The goal is to arrive at a homogeneous blend of two solutions in as little time as possible. We also stressed that passive mixing on microdevices depends solely on diffusion as the transport mechanism. All strategies to improve mixing have therefore looked into the possibilities of either reducing the diffusion distances or inducing some kind of extra movement (typically lateral to the flow direction). Alternatively, active mixers, in which external energy is used to induce turbulence, have been investigated as well. Examples of how mixers are implemented on microsystems are given in this section.

The Einstein–Smoluchowski equation describes the dependence between the diffusion distance and the diffusion time. If we reduce the diffusion distance by a factor of 2, the diffusion time is reduced by a factor of 4; if the distance is reduced 10-fold, the time is reduced 100-fold. To obtain short diffusion distances we could design narrow or high-aspect ratio channels to begin with. However, just narrowing chan-

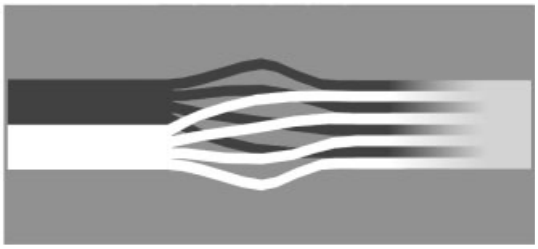


Fig. 4.11 Principle of a laminating mixer – the main idea is to minimize the necessary diffusion lengths.

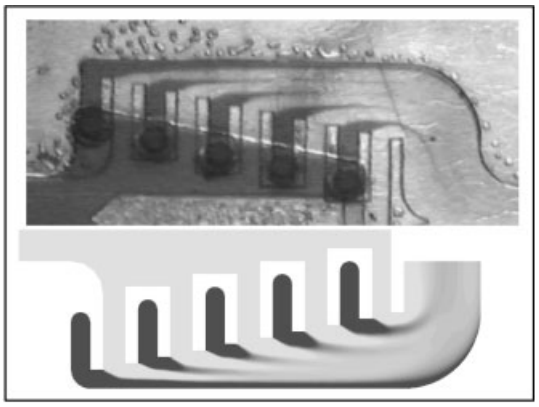


Fig. 4.12 Experiment and simulation of a laminating mixing structure (courtesy U. D. Larsen, Chempaq [36]).

nels results in higher fluidic resistance and may therefore not be a good choice. Channels of high aspect ratio (narrow and deep), on the other hand, are harder to fabricate and require special machining technology. One possible strategy is to work with standard-sized channels, split each channel into an *array* of smaller channels (thereby circumventing the high fluidic resistance of just *one* small channel), and then merge them again in such a way that the split flows of solution A get interlaced with the split flows of solution B (Fig. 4.11). Although it looks simple enough in this schematic picture, it is actually quite difficult to implement, because it requires adding a third dimension to the typically planar, quasi two-dimensional, microliquid device. One of the two types of solutions has to be brought in from above or below to properly interlace with the other solution. Fig. 4.12 shows a microphotograph of a mixer implemented in such a fashion, together with results from a simulation of the same device. An indicator solution (dark) is brought in from below through an arrangement of openings to be interlaced with an acid solution (clear), resulting in a grey color. From the dark-colored streaks we can see that it takes some time for diffusion to mix both solutions completely. This process takes much longer at the walls, because there the indicator is only touched on one side by the acid. The simulations nicely match the experimental results. One of the main disadvantages of such a mixer design is its large volume. If two small well-defined plugs of chemicals are to be mixed in such a mixer, the excess volume might lead to strong dispersion

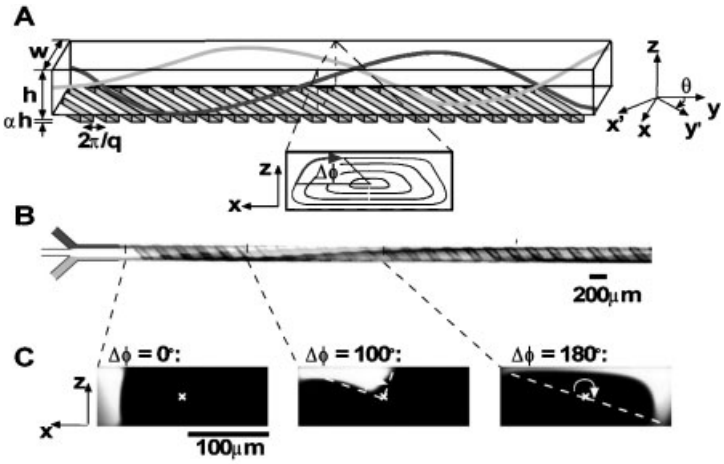


Fig. 4.13 Fluidic streamlines obtained with obliquely grooved surfaces: A. schematic streamlines; B. experiment (top-view); C: cross-sectional views with one fluorescing fluid (reprinted with permission from [27]. Copyright (2002) American Association for the Advancement of Science).

and hence to dilution of the combined plug after mixing, which in turn could slow down or prohibit a reaction or result in a low signal-to-noise ratio for detection.

As mentioned in Chapter 3, the Peclét number (*Pe*) is a good indicator of whether diffusive mixing alone is sufficient, given a certain channel width (i.e., a certain minimum diffusion length) and a certain linear flow velocity. The higher the Peclét number the harder it is to achieve mixing by diffusion, especially within a limited amount of channel length. In fact, the necessary channel length increases linearly with the Peclét number [27]. For this reason, ways to induce a lateral transport component (i.e., an additional flow vector more or less orthogonal to the main flow direction) to complement diffusion and improve mixing are necessary. One such possibility is to machine grooves into the channel bottom, where these grooves are of certain depths, arranged at a certain angle with respect to the main flow direction, and arranged in a number of patterns [27, 28]. For pressure-driven flow, these grooves introduce anisotropic flow resistances into an otherwise isotropic system. Fluids experience less resistance when flowing along the ridges and valleys constituting the grooves than when flowing perpendicular to them. In this way, a rotational element is introduced, resulting in corkscrew-like fluidic streamlines (Fig. 4.13). A careful arrangement of different types of grooves (mainly differing with respect to their orientation) can produce chaotic stirring, which allows mixing in much less time and shorter channel length [27]. Similar ideas were also tested with electroosmotic flow. Here, due to the oblique arrangement of the grooves there is a tiny electrical field drop along the grooves, inducing an electroosmotic flow component that is lateral to the main flow direction. Experiments showed that it was again not enough to use only one kind of groove. Instead, alternating grooves with different orientations greatly improved the mixing efficiency (Fig. 4.14) [28].

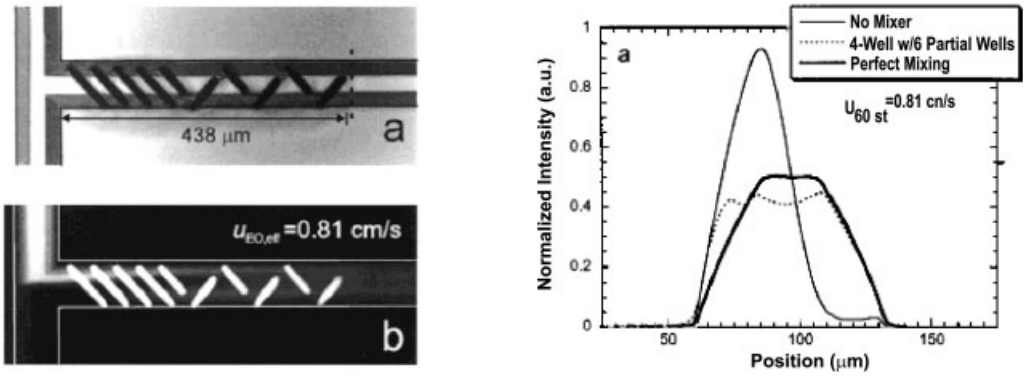


Fig. 4.14 Electroosmotically-driven mixer with grooved wells micromachined into the channel surface (reprinted with permission from [28]. Copyright (2002) American Chemical Society).

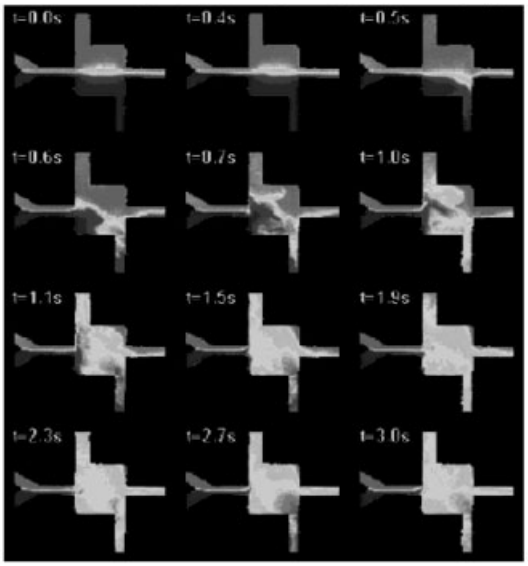


Fig. 4.15 Sequence of mixing using electrokinetically induced flow instabilities; from left to right are shown the Y-junction inlet, the main mixing chamber with two channels connecting to the AC electrodes, and the common exit channel (reprinted with permission from [29]. Copyright (2001) American Chemical Society).

All the above examples are passive mixers, relying solely on energy from within the system to facilitate mixing. Active mixers can also be designed, in which external energy in some form induces chaotic behavior, leading to improved, accelerated mixing. One example is shown in Fig. 4.15, where a sinusoidally oscillating electric field is used to induce flow instabilities, which in turn lead to stretching

and folding of the fluids and hence mixing. In the design used, a mixing chamber has two electrodes connected to it, where the oscillating electric field can be applied. Given the right set of parameters, mainly with respect to applied field strength and oscillating frequency, flow instabilities can be observed manifesting themselves as seemingly random and transverse velocity fluctuations. Again, these additional velocity components lead to chaotic behavior, resulting in enhanced mixing within a shorter timeframe [29].

4.2 Injecting, Dosing, and Metering

A very important fluidic handling function is the ability to dispense very defined (small) amounts or volumes of solutions, repeatedly, on demand, and very reproducibly. 'Defined' in this context can refer either to a volume whose size is not exactly known but which is always the same or to a volume whose size is exactly known and given by, e.g., a particular geometric arrangement of channel segments. The latter is also known as 'metering'. For example, in the microfluidic structure in Fig. 4.7, liquid was introduced into the channels through an opening hole, and all available channels were filled by capillary action. Fluid flow was stopped by burst valves based on hydrophobic patches. After filling, the CD was set into motion, inducing flow where it was not prohibited by the hydrophobic patches. In this way, fluid was removed from excess channels leaving a volume of fluid behind, as defined by the channel geometry between the excess channel and the hydrophobic valves. Thus, these volumes were metered off. Any similar concerted action of pumping and valving involving geometrically defined volumes can in principle be used to meter off defined amounts of solutions.

For many analytical techniques it is beneficial to be able to inject small, well-defined amounts of sample solutions (see also chapter 10). This is also referred to as aliquoting. Popular implementations of an aliquoting or injection function on microchips make use of a channel cross or a double T arrangement along with electrokinetic fluid manipulation to achieve this. Let's have a closer look at what is going on at a simple channel cross (Fig. 4.16, left). A popular approach to describing the electrical behavior of a fluidic channel network is to use electrical equivalent circuits, e.g., representing the channel cross by several nodes and resistors (Fig. 4.16, right). There are five nodes – four at the terminal ends and one at the intersection. Connecting the nodes are four resistors representing the impedance in the four channel segments. The only external control we have is the voltage we apply at the four terminal nodes, whereas we can only calculate the voltage at the inner node, once we have measured and established the individual impedances. Now, with a few relatively simple relations, such as Ohm's Law and the Kirchhoff rules, we can calculate which voltages need to be applied in order to have current following a specific pathway through the channel network. And, assuming similitude between the electric field and the fluidic flow, transport of bulk liquid and chemical species should also follow this pattern.

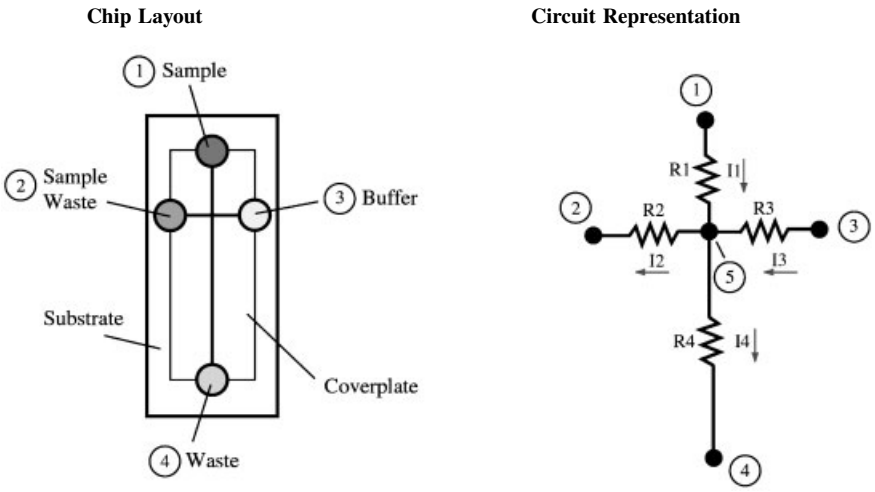


Fig. 4.16 Schematic chip layout with a simple cross for injections (left) and the equivalent electrical circuit representation (right).

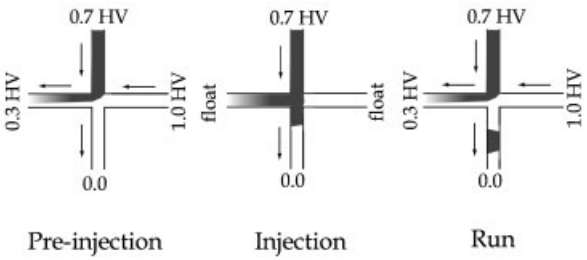


Fig. 4.17 Principle of gated injection.

Fig. 4.17 illustrates the application of these ideas to the implementation of a so-called gated (electrokinetic) valve or injector [30, 31]. The main waste reservoir at the end of the main channel is set to ground. Thus, it functions as the anode to which the electroosmotic flow is directed. All other reservoirs are at higher positive potentials. The topmost reservoir contains the sample and, to ensure pumping out of the reservoir, the potential is set to a somewhat large value, indicated here as a percent fraction of a specific limit value. For the same reason, the sample waste reservoir is set to a relatively low voltage to receive the sample flow. The remaining reservoir is filled with buffer and is set at a slightly higher voltage than the sample reservoir. This allows flow out of this reservoir, to keep sample from entering (bleeding into) the main channel by flowing right across the junction and into the sample waste reservoir. At the same time, this reservoir also provides flow of fresh buffer into the main channel. The distribution of voltages described so far corresponds

to the situation before an injection. Depending on the specifications of the power supplies used, the simplest way to facilitate injection is to remove the power supplies at the sample waste and the buffer reservoir from the electric circuit. This brings their potential to exactly the same potential as at the intersection at this moment, thereby preventing any flow to or from the side arms. Voltage is still applied at the other reservoirs, allowing sample to flow into the main channel. The injection step is finished as soon as the original voltage distribution is established again, thereby cutting off the sample flow into the main channel and releasing a plug of sample flowing towards the anode. Because the injected volume is mainly determined by the timing of the injection sequence and sample can enter the main channel for as long as this electrokinetic valve is open, this injection procedure is called 'gated injection'. There is one caveat, however: because electrokinetic forces (i. e., electroosmosis and electrophoresis) are used to inject the sample, the injection is biased, meaning that during the time of injection, positively charged ions are injected to a larger extent than neutral species, and neutral species to a larger extent than negatively charged ions. This is, of course, because the cations have an electrophoretic velocity in the same direction as the electroosmotic flow, while neutrals have no electrophoretic velocity, and anions have an electrophoretic velocity opposed to the electroosmotic flow. This fact needs be taken into consideration when designing a method for quantitative analysis involving one or more gated injection steps.

There is an alternative to the gated injection method, which also utilizes the cross layout already mentioned. Here, the sample reservoir is on one of the sidearms, and prior to injection, sample is continuously pumped across the intersection to the waste reservoir (Fig. 4.18). Additionally, the voltages are arranged so that there is also flow from the buffer reservoir and the reservoir at the end of the main channel towards the intersection and on to the sample waste reservoir. This is again to avoid premature bleeding of sample into the main channel. By adjusting the voltages (i.e., the flow rates) one can pinch the sample stream more or less on its passage through the intersection (see also Fig. 3.6 on flow focusing). Therefore, this technique is often called 'pinched injection' [30, 31]. The main differences with respect to gated injection are:

- The injection volume is pre-defined and fixed. Only the volume defined by the junction geometry and the pinching is injected once the voltages are switched accordingly (Fig. 4.18).

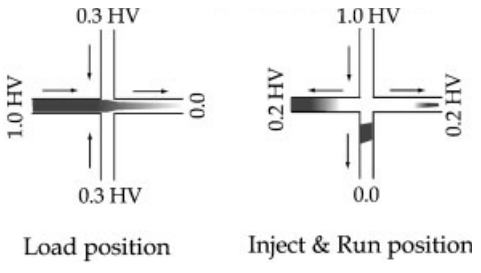


Fig. 4.18 Principle of pinched injection.

- If the loading step is given sufficient time, then the composition of the sample solution present at the intersection is the same as the composition of the original sample solution, and consequently there is no electrokinetic bias upon injection.

A variation of the pinched injection is the so-called double-T injection, where the two side-arms of the cross geometry are offset by a certain distance, thereby increasing the volume to be injected. Although this volume is larger than in the original pinched injection, it is again fixed and cannot be changed once the chip is fabricated. (Of course, various degrees of pinching slightly alter the injected volume). Other variants have also been investigated, in which either geometries (channel dimensions) or voltage switching protocols have been altered [32, 33]. These injection procedures are all based on electrokinetic flow control and it is relatively cumbersome to implement similar schemes with pressure-driven flows. However, it is also possible to design a gated injection valve with a field-free injection region using electroosmotically induced hydrodynamic flow.

Still, dosing is of course also possible with other means than electrokinetic fluid handling. Often, dosing of a defined volume can be achieved with pumps or pump-like microstructures, in which a single stroke of, e.g., a membrane, displaces a certain amount of fluid out of a pump chamber, through a valve or just through some orifice. Well-known implementations of such micromechanical dosing systems are ink-jet printer cartridges. In these cartridges the actuation can be achieved by a gas bubble, which is initiated on top of a heating element, causing displacement of fluid through an opening as the bubble grows. When the current is switched off the bubble shrinks again, allowing fresh material to enter the reservoir for the next dosing action. Alternatively, piezoelectric actuators are often used, actuating a membrane to push on an enclosed volume of fluid in order to press part of it out of an opening. Piezoactuators can be operated at high frequencies, allowing high droplet generation rates. An example of a piezoelectrically operated microdispenser is shown in Fig. 4.19 [34]. Droplets generated from such systems have also been used in chemical settings, e.g., in connection with acoustically levitated droplets, which constitute wall-less chemical containers and which can be loaded with tiny droplets from piezodispensers to allow many chemical reactions to take place within the large levitated droplet [35].

In general, dosing with microsystems is considered to be of great importance in the development of medical microdevices, which not only diagnose and monitor certain body functions, but, at the same time, also can administer drugs in precise amounts to regulate those functions. Such systems will allow patients to enjoy more independence from stationary medical treatment and probably also reduce the overall amount of drugs necessary for a treatment.

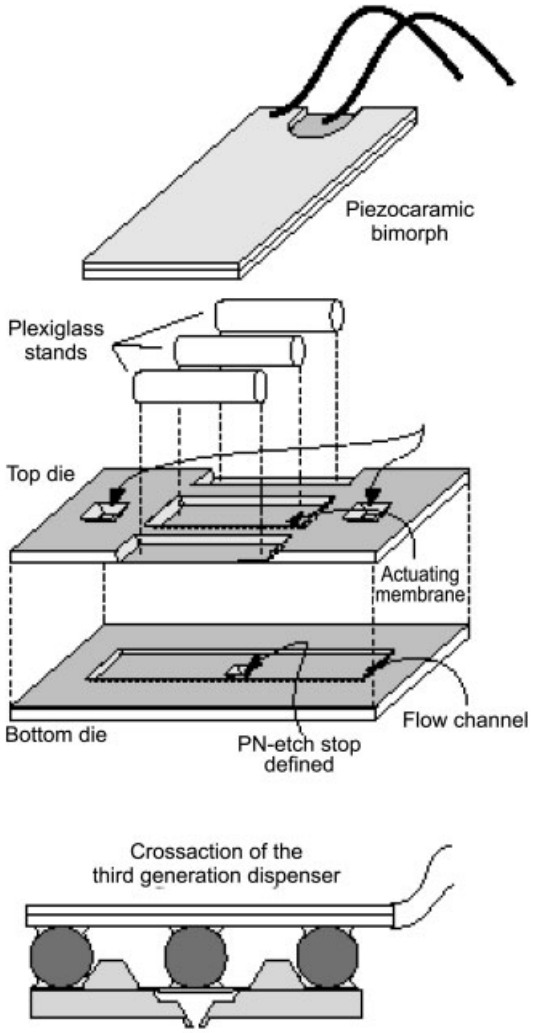


Fig. 4.19 Schematic of a dispensing microsystem using piezoactuation (reprinted with permisssion from [34]. Copyright (1999) Institute of Physics).

4.3
Temperature Measurement in Microfluidic Systems

Apart from precise dosing of fluids in a microsystems, it is often desirable and sometimes necessary to measure the temperature. Microsystems can be used for production or analysis of biological or chemical components. In both cases, the temperature dependence of chemical and biological reactions makes temperature measurement desirable or even necessary.

When a microfluidic system is used as an analyzer, the system is concerned with measuring the concentration of biological or chemical species; this measurement is almost always directly or indirectly temperature dependent. Most often,

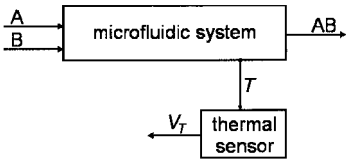


Fig. 4.20 A microfluidic system that is used as an analyzer. Similar to the reaction in Eq. 4.4, the input reagents are a sample A and an additional reagent B, the product AB of the reaction is the waste output of the system to the right.

measurement of concentration relies on a biological or chemical reaction that occurs within the microsystem. Such a reaction can be endothermic (taking up heat from the surroundings) or exothermic (releasing heat into the surroundings), and the reaction speed is also temperature dependent. The reaction can typically be described as

$$\mathrm{A} + \mathrm{B} \xrightarrow{k} \mathrm{AB} + \mathrm{Q}\,, \tag{4.4}$$

where A and B are chemical reactants, AB the reaction product, k the reaction constant, and Q the generated heat. Reaction (4.4) described here is exothermic, because Q was chosen to be positive.

An important observation is that the rate constant k is temperature dependent. This means that if the temperature changes by any means, whether by external influences or by the reaction heat Q itself, the measured results also change. At the wrong temperature, the reaction might occur too late in the microsystem, so that the reagents get flushed out of the system, or the reaction might occur too early. Either way, the target species concentration is not measured accurately.

To monitor the reaction, the use of a temperature sensor is advisable (Fig. 4.20). Measuring the temperature is also the first step needed for active control of the reaction or for temperature compensation, if control or compensation is needed.

4.3.1 Microreactors

Chemical and biochemical microsystems can also be used for production of components; here, the system is called a microreactor. Temperature sensors play an important role in such microreactors, because chemical reactions are usually strongly temperature dependent, as discussed above. A relatively simple microreactor is sketched in Fig. 4.21.

The reactor has incoming reagents A and B, and product AB leaves the reactor. The temperature T in the microreactor can be monitored with the thermal sensor, which outputs a signal V_T that corresponds to T. The microreactor is also equipped with a heater/cooler unit to make it possible to keep the reactor at a specified temperature. Together with the thermal sensor, the heater/cooler unit forms a closed-loop controller for the microreactor. Such a controller can be very useful, for instance for polymerase chain reaction (PCR), where a microreaction chamber cycles through several different temperatures in sequence.

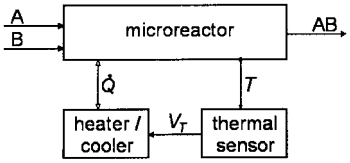


Fig. 4.21 A microreactor with two input reagents A and B, and output product AB. A heater/cooler unit removes heat $\dot{Q}$ from the microreactor. A thermal sensor measures the reactor temperature T and supplies this information to the heater/cooler unit via its output signal V_T.

If the temperature change is small and negligible for the reaction under consideration, then the thermal sensor is not needed. If however the temperature change is considerable, then the change must be recorded for further processing or used for direct temperature compensation. Finally, the temperature change itself can be the signal of interest, when the reaction heat is measured to find out how much of the product was produced. In this case, it is important that no extra heat enters or leaves the microreactor, so that the rise in temperature is proportional to the internally produced heat. If a microsystem is used this way, it is called a microcalorimeter.

Apart from the interest in measuring temperature as a prime goal, temperature measurement is also important in other areas, especially for flow measurement.

4.3.2
Temperature Sensors for Microsystems

Of the many different types of temperature measuring principles, only a few are of principal interest for chemical and biochemical microsystems. These main sensing principles are measuring a change in resistance of a material, measuring a thermoelectric voltage, and exploiting the thermal dependence of electronic parameters of *p-n*-junctions. Which sensor is used in a given application depends in general on the needed sensitivity, resolution, range, bandwidth, signal-to-noise ratio, cost, and last but not least, the size of the sensor. Usually a sensor for a chemical or biochemical microsystem is as small as possible. Small size has the added advantage of a small heat capacity and thus a relatively fast response of the sensor. Finally, it is desirable to be able to integrate the fabrication of the sensor into the production process of the microsystem. Sensitivity, resolution, range, signal-to-noise ratio, and cost depend on the individually chosen sensors, which are discussed below.

4.3.3
Resistance Temperature Detectors

4.3.3.1 Metals

If the change in the resistance of a metal with temperature is measured, the resulting sensor is known as a resistance temperature detector (RTD). Platinum is the metal of choice, because it is corrosion resistant and stable over a wide temperature range. Platinum resistance sensors are the industry standard, due to their accuracy and their reproducibility [37]. Within a temperature range of 15 to

725 K the accuracy of platinum resistance sensors can be as low as 0.01 K [1]. Resistance sensors can be integrated into a microfluidic sensor with comparative ease by depositing thin films. Not only platinum is used for microsystems but also other metals can be deposited, for example, a nickel–iron alloy [38].

Platinum resistance sensors are ubiquitous in microfluidic systems for polymerase chain reaction (PCR), where a certain temperature–time profile has to be driven in a miniature vessel. The resistance sensor is often deposited between integrated heaters, so that the sensor together with the heaters form part of a closed-loop heat control. A typical example of such a PCR chamber with temperature controller based on platinum resistors is the microsystem shown in Fig. 4.22 and Fig. 4.23.

An example of a platinum resistance sensor is one used by Zhan *et al.* [39]. They present a microfluidic system for PCR, in which both the heaters and the sensor were made from a deposited and then patterned thin film of platinum. The temperatures used in the microvessel were slightly below 100 °C, and the accuracy achieved in the system was 0.5 K.

4.3.3.2 **Nonmetals**

Resistance temperature sensors can also be produced from nonmetals. Ceramics and oxides on the one hand and semiconductors on the other hand are especially interesting, for different reasons.

A disadvantage of the platinum resistance sensor is its relatively low sensitivity, less than 1% K^{-1}. Some ceramics and oxides offer much greater sensitivity, 4%–6% K^{-1}, which makes them attractive as temperature sensors, because larger sig-

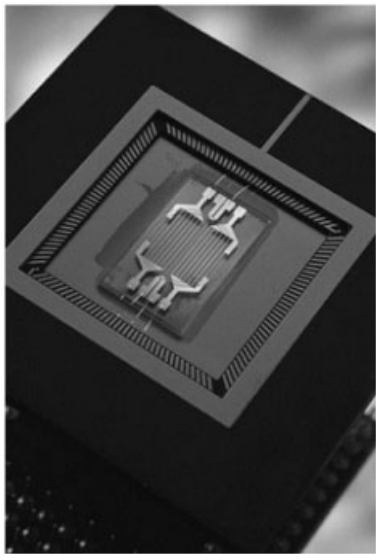


Fig. 4.22 PCR chip. A microsystem with integrated resistive heater and resistance temperature detector (RTD) is shown from above. The RTD is the single central wire, flanked by several heating wires to the left and the right. The wires are deposited platinum. A cross section is shown in Fig. 4.23 (courtesy El-Ali and Wolff, MIC).

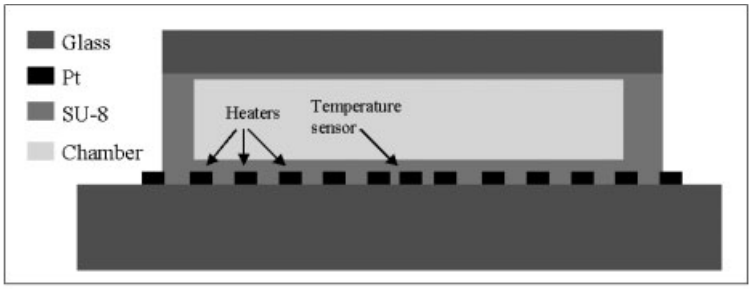


Fig. 4.23 Chamber cross section. The PCR microsystem of Fig. 4.22 is depicted schematically in cross section (courtesy El-Ali and Wolff, MIC).

nals for the same temperature change mean a higher signal-to-noise ratio. These nonmetal resistance temperature sensors are known as thermistors. There are two major types, one having a negative temperature coefficient (NTC thermistor) and the other having a positive temperature coefficient (PTC thermistor). The drawback of thermistors is their strong nonlinear temperature dependence, which calls for more intensive signal processing. Thermistors have an interchangeable temperature tolerance of 0.1 K, although the accuracy can be increased by individually calibrating the sensors.

Using silicon as a sensor material has the advantage that the sensor can be easily integrated into a silicon microfluidic system. Hauptmann [40] describes a silicon resistance temperature sensor based on the principle of spreading resistance. This sensor measures the resistance of silicon between a flat back contact and a point front contact. The dependence on temperature is slightly nonlinear, although not as drastically as for a thermistor. The temperature coefficient or the sensitivity of the sensor is 0.8% K^{-1}, which is lower than that for thermistors, but still higher than that for platinum. Such sensors can be acquired commercially as discrete devices from Philips Semiconductors, Eindhoven, The Netherlands; examples being devices from the KTY84-1 series.

Resistance temperature detectors are usually measured by a four-wire technique. To measure resistance, the minimum number of connections to a resistor is two, one at each terminal. A voltage is then established across the resistor and the flowing current is measured, which allows the resistance to be calculated with Ohm's law. A systematic error, however, is introduced in this measurement setup, which is due to the additional connection cable length. This systematic error can be avoided by using a four-wire measurement technique. Here, there are two connection cables on each terminal. A constant current is sent through one set of cables, while the resulting voltage drop is measured with the other set. The four-wire technique is superior to the two-wire technique, because the connection wire resistance is much lower than the input resistance of a voltage meter, and voltage is measured instead of current.

4.3.4 Thermocouples

Thermocouples are very popular temperature sensors, which rely on a thermoelectric phenomenon, the Seebeck effect. The essential parts of a thermocouple are two wires of different metals. If the wires are welded together at one end and a temperature difference is applied between this point and the other two ends, then a small voltage proportional to the temperature difference is generated. This is known as the Seebeck effect. In practice, the other two ends are also welded together, and one of the wires is cut open for voltage measurement. This procedure prevents additional thermovoltages from forming between the thermocouple ends and the voltage measurement leads.

Different metal or metal-alloy combinations are useful for thermocouples. Some combinations, such as chromel–constantan and iron–constantan, optimize the sensitivity of the thermocouple, but others, such as chromel–alumel copper–constantan, and platinum/rhodium–platinum, make it possible for the thermocouple to operate over an especially large range or within an unusually cold or hot temperature region.

In comparison to a resistance temperature sensor, thermocouples have the advantage that they work over a very large temperature range, of up to 3000 K (Kovacs 1998). Thermocouples are especially suitable when a low mass is needed and temperature differences are to be measured. The various thermocouple material combinations result in different values of sensitivity and accuracy. The platinum/rhodium–platinum thermocouple, for instance, has a sensitivity of only 6 $\mu V\ K^{-1}$ with a relative accuracy of 0.25%. The chromel–alumel thermocouple has a tenfold sensitivity of 60 $\mu V\ K^{-1}$, also accompanied by a higher relative accuracy of 1% (van Putten 1996). The absolute accuracy of a calibrated thermocouple is in the range of 0.5 to 2 K.

Thermocouples can also be produced from moderately doped semiconductor material (Kovacs 1998), which makes it comparatively easy to integrate them in a microfluidic chip.

4.3.5 Semiconductor Junction Sensors

Silicon, which has long been the most abundant material used for the micromachining of microfluidic devices, can be used as a temperature sensor in more than one way. A silicon resistance sensor was mentioned above; here, exploitation of the temperature dependence of one or more *p*–*n* junctions is discussed. Several electronic variables of a *p*–*n* junction depend on temperature; the saturation current of a diode and the base-emitter voltage of a transistor are examples.

A diode can be used as a temperature sensor in a similar fashion as a resistance temperature sensor. A certain constant current is sent through the diode, and the voltage across is measured with a voltmeter (Kovacs 1998, Yeager and Courts 2001). The sensitivity of such a temperature-measuring diode is 2.0–

2.4 mV K^{-1} (Kovacs 1998). The range of the diode sensor is from 1–500 K, and diode sensors can achieve an accuracy of 0.1 K, when calibrated.

Using the temperature dependence of the base-emitter voltage of a transistor makes it possible to design a compensation circuit that produces a sensor signal that is proportional to the absolute temperature. The base-emitter voltage (V_{BE}) is given by

$$V_{BE} = \frac{kT}{q} \ln\left(\frac{I_C}{I_S}\right) \tag{4.5}$$

where k is Boltzmann's constant, T is the temperature, q is the elementary charge, I_C is the collector current, and I_S the saturation current. The base-emitter voltage itself depends on temperature in a rather complicated way, mostly because the saturation current (I_s) depends on several temperature-dependent variables, such as the intrinsic carrier density, the diffusion coefficients, and the diffusion length. However, by looking at the difference between the base-emitter voltages from two transistors having different collector currents, the temperature dependence of the saturation current can be eliminated. The voltage difference is found to be

$$\Delta V = V_{BE1} - B_{BE2} = \frac{kT}{q} \ln\left(\frac{I_{C1}}{I_{C2}} \frac{I_{S2}}{I_{S1}}\right) \tag{4.6}$$

which leaves the resulting circuit output signal (ΔV) proportional to the absolute temperature. Such a circuit is therefore often abbreviated PTAT.

Such a PTAT circuit can have an accuracy of 1 K at a sensitivity of 10 mV K^{-1} within a range of –25 °C to 85 °C (about 250 to 360 K). Discrete versions of this circuit are also available commercially, for instance, as part number AD549 from Analog Devices, Norwood, MA, USA. Frank [41] describes how PTAT temperature sensors are embedded into a larger integrated circuit to provide means for detecting whether temperature, voltage, or current are out of limits. The overall product is a so-called smart power integrated circuit. It is also possible to integrate a diode or a PTAT circuit into a microfluidic silicon system.

4.3.6 Temperature Sensors Built on Other Principles

In addition to the widespread and well-known temperature sensors mentioned so far, devises based on other principles are also used to measure temperature in microfluidic systems.

Fluorescence thermometry [42, 43] and liquid-crystal thermometry allow spatially resolved fluid measurements. Both methods are based on the distribution of temperature-sensitive elements all over the microfluidic system. When liquid crystals are used as elements, different colors of the crystals indicate different temperatures. When fluorescence is used, the typical fluorescent decay time indicates the temperature. Chaudhari *et al.* [44] reported temperature measurements in a

PCR vessel using liquid crystals. They achieved a measurement accuracy of 0.1 K within two rather small temperature intervals of 54.5–55.5 °C and 94–95 °C.

Acoustic waves in solids can also be used for temperature sensing [45]. Surface acoustic waves and plate waves are mainly used, where the waves travel in a material interfacing the fluid in the microsystem. The waves can be detected electronically, because the material the waves travel in is piezoelectric. The output signal of the devices is a frequency, which makes it convenient for electronic signal processing. The frequencies are in the megahertz range. Although the sensitivity achieved is relatively low at a maximum value of about 200 ppm K^{-1}, the accuracy is relatively high at about 0.2 K.

Temperatures can also be measured mechanically by using a bimetallic strip. Such a strip consists of two materials with different thermal expansion coefficients, so that the strip bends when the temperature changes. A bimetallic strip is usually used as a thermally activated switch (Kovacs 1998). However, the strip can also be used to measure temperature continuously by measuring the extent of bending with a strain gauge. Marie et al. [46] used micromachined cantilevers for temperature control in a PCR chamber, with the advantage of having the temperature sensor directly immersed in the fluid. The second advantage is the convenience of being able to use another cantilever in addition to the ones already in use for a different purpose.

4.3.7 Conclusion

If it is necessary to measure temperature in a microfluidic system, then the choice of the thermal sensor should be based mainly on whether it is a semiconductor-based microsystem, in what range the temperature has to be measured, and to what accuracy, while keeping cost in mind.

The temperature sensors most often used in microfluidic devices are platinum resistance temperature detectors and thermocouples, although other sensors, such as thermistors and diodes, might be increasingly used for mass-produced devices.

Depending on the microsystem under consideration, it might be convenient to choose a sensor principle that is easily integrated into the system, such as a silicon diode in a silicon microsystem or a bimetallic strip in a cantilever-based measurement system.

4.4 Optical Sensors

Optical sensors also play an important role in microsystems. We all use visual inspection of liquids in our everyday life, e.g., estimating the strength of coffee from its color. Coffee appears dark, because light is absorbed by the liquid, so a very dark color corresponds to a high concentration of absorbing molecules. This is one of many principles (absorbance detection) that are used for optical detection in chemical analysis systems. A wide range of other optical phenomena are used to relate a

concentration to an optical signal. Among these, luminescence is important: a molecule is excited to a higher energy state and relaxes to the ground state by emission of optical energy. The optical power of the emission can be related to the concentration of molecules and hence can be used for detection. The various optical measurement schemes are discussed in sections 4.4.2, 4.4.3, and 4.4.4, after a short introduction to the instrumentation that is necessary in a complete chemical analysis system.

4.4.1 Instrumentation

In most microanalytical systems neither the light source nor the photodetector is integrated into the same substrate as the fluidic channel network. The reason is that monolithic integration of all components necessary for performing a total chemical analysis is very complex. Thus, most research groups have focused on a few aspects of a whole system, such as the fluidics, in which a typical system consists of an assembly of different modules for which efficient interconnections are very important. The most widely used optical interconnection methods are optical fibers and free-space optical elements such as lenses, mirrors, and filters to guide light into the channel network and also to collect optical signals from the same fluidic channels. Typical modules are outlined in Fig. 4.24. Examples of devices are given in the following sections, where optical elements such as planar waveguides and photodiodes are integrated with a microfluidic channel network.

A light source is often needed (not necessary for chemiluminescence measurements), and a very popular choice is the laser (light amplification by stimulated emission of radiation), because it has a high optical power and is monochromatic, which means that the light is confined to a very narrow wavelength range. A major limitation

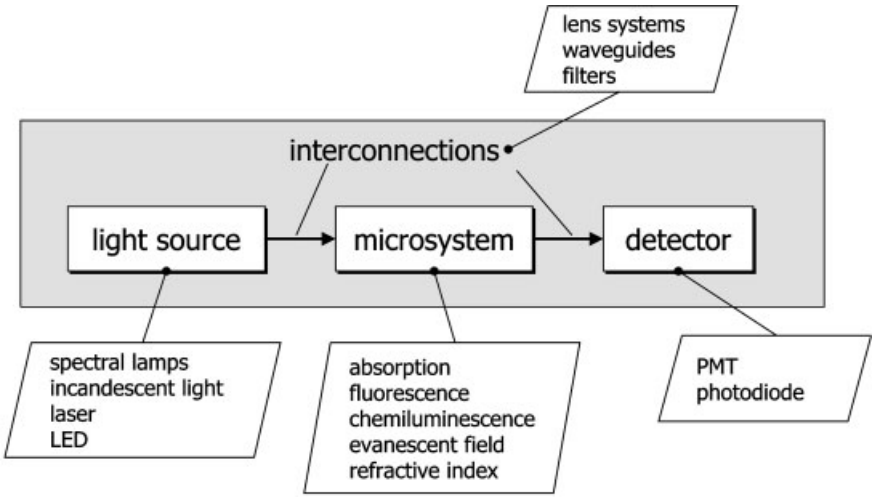


Fig. 4.24 Flowchart of the basic units of a typical microfluidic system based on optical detection.

of lasers is that they are often large and expensive. This is especially true for lasers operating in the ultraviolet range (UV), which is about 200–400 nm. In this range other light sources such as spectral lamps and incandescent lamps are used. Another popular choice is the light-emitting diode (LED), because of its very small size (a few millimeters on each side). Its smaller optical output power is often a limitation.

On the detector side the two most popular types are the photomultiplier tube (PMT) and the semiconductor photodiode. The photomultiplier tube internally amplifies the detected signal, which is the reason for the name 'photomultiplier': the incoming signal is optically 'multiplied', which enables measurements at low light levels.

4.4.2
Absorption Detection

Absorption of light is one of the most fundamental aspects used in optical detection and is thus very important to be familiar with. Absorption means that optical energy (photons) is converted into another type of energy, e.g., into thermal energy in the absorbing solution – which results in an increase in the temperature of the solution. In the ultraviolet and visible wavelength range (180–700 nm) light absorption is typically due to an electronic transition to a higher energy state (Fig. 4.25). (In fluorescence, section 4.4.4, a fraction of the optical energy is not converted into heat but is re-emitted as photons having lower energy and hence longer wavelength.)

Fig. 4.25 (left) shows a schematic drawing of an absorption measurement. An absorbing solution with a concentration c is contained in a transparent cuvette, and light of a fixed wavelength is passed through it. The so-called absorbance value A can be calculated from the transmittance, $T(c)$, which is the ratio between the optical power with light attenuation due to absorption by the solution and without attenuation due to absorption. This value (which is of course smaller than 1) can be estimated from the ratio between $P(c)$ and P_0, as seen in Fig. 4.25 (left). This is, however, a poor estimate, because contributions in addition to absorption by the so-

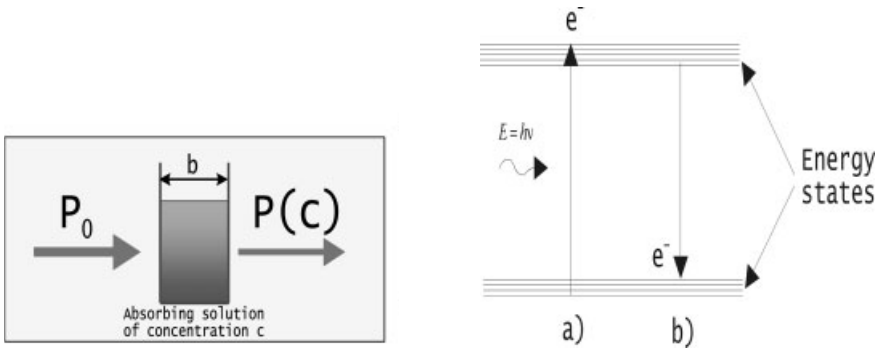


Fig. 4.25 Left: Absorption experiment in which light passes through an optical cuvette of length b. Right: Absorption process by which optical energy ($E = h\nu$) is converted into thermal energy. a) Radiative excitation. b) Radiation-less decay.

lution also decrease the transmitted optical power. For example, reflections at interfaces of different refractive indices reduce the transmitted optical power. To avoid this problem, a reference measurement is generally made with a transparent solution in the cuvette P_{solvent}. This is typically the solvent that was used to prepare the absorbing solution. Hence, the transmittance is given by

$$T(c) = \frac{P(c)}{P_{\text{solvent}}} > \frac{P(c)}{P_0} \tag{4.7}$$

The absorbance value A is calculated with the relation:

$$A(c) = -\log[T(c)] = \log\left[\frac{P_{\text{solvent}}}{P(c)}\right] \tag{4.8}$$

and the absorbance is seen to increase with a decrease in the transmitted optical power. A very important formula is the Lambert–Beer law, which states that the calculated absorbance is proportional to the concentration of the absorbing solution and to the optical path length b, used during the measurements.

$$A(c) = \varepsilon bc \tag{4.9}$$

ε is the molar absorptivity. This value is a property of the chemical species and also depends on the wavelength. It is thus necessary to choose a wavelength with a high value of ε to obtain the greatest signal for a given concentration and optical path length. The Lambert–Beer law is generally valid for low concentrations and monochromatic light.

Eq. 4.9 indicates that an increase in the optical path length is desirable, because it leads to an increase in the calculated absorbance value and allows measurement of lower concentrations c. The sensitivity of absorbance measurements thus scales poorly with miniaturization of the channel dimensions, because it often results in a decrease in the optical path length. Fig. 4.26 shows a possible way of overcoming problems with a short light path length.

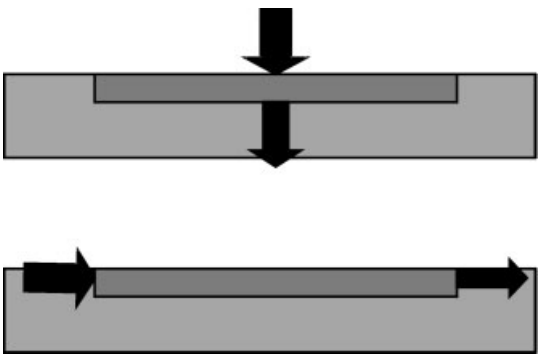


Fig. 4.26 Top: Absorbance measurement perpendicular to the plane of the microdevice, resulting in a short path length. Bottom: Absorbance measurement in the plane of the microdevice, resulting in an increased optical path length.

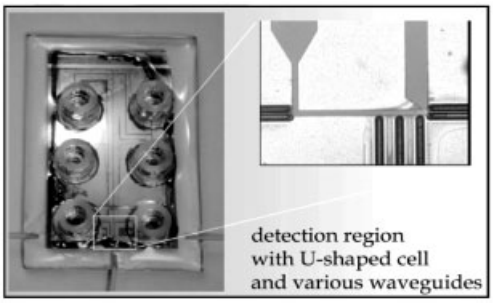


Fig. 4.27 Photograph of a capillary electrophoresis device with integrated planar waveguides for absorption detection. The inset picture shows a 750-μm-long absorption cell [47].

If the measurement can be done in the plane of the device, a longer path length is possible, because it is not restricted to the depth of the channel, which is ultimately limited by the substrate thickness. In a practical application the channel depth is typically less than 100 μm, because a low volume is desired. This principle was used in a device in which optical waveguides were integrated with microfluidic channels to guide the light in the plane of the chip (Fig. 4.27). Planar waveguides are transparent structures that consist of a material of high refractive index surrounded by a material with a lower refractive index. Guidance of light is thus obtained by total internal reflection, which is also used in optical fibers.

A photograph of the device (Fig. 4.27, left) shows the six reservoirs that contain the chemicals during analysis at the top of the device. Electroosmotic pumping and separation by means of capillary electrophoresis can be performed by switching the voltage between electrodes that are placed in the reservoirs (see section 4.1). The inset picture on the right shows a section of the channels that is interrogated by two planar waveguides. This configuration increases the optical path length by enabling absorbance to be measured along the channel. The absorption path length is 750 μm, which is more than 50 times the channel depth (13 μm). The left picture also shows how light is guided into the integrated planar waveguides by optical fibers glued onto the endface of the chip. Such a device can in principle be connected to the light source and photodetector in a 'plug-and-play' manner.

In a similar approach, optical fibers were inserted into etched grooves and used to illuminate and collect light from a small channel segment. Here, integration of planar waveguides is avoided, but optical fibers allow detection at only a single point per fiber, in contrast to planar waveguides, which can be branched into many detection points.

In both approaches the light beam passes unguided through the detection channel, and a major fraction of the light is typically lost, because it is not collected by the waveguide or optical fiber on the other side of the channel. It propagates as stray light in the plane of the device or radiates out of the chip. This problem has been addressed by fabrication of so-called evanescent-wave sensors, in which the beam is also guided in the detection region.

4.4.3 Evanescent-wave Sensing

In a planar waveguide or in an optical fiber, all the optical power is not confined in the core layer, but a fraction of the electromagnetic field decays exponentially to zero outside the waveguide core region. This area is called the evanescent field region (Fig. 4.28) and can be several micrometers thick, depending on the refractive index of the various layers and the wavelength of the light.

Normally, a so-called waveguide-cladding layer with a low refractive index is deposited on top of the high-index waveguide core layer to prevent interference from the external environment. But for chemical microsystems it is possible to take advantage of this interference by using it for sensing. This principle has been used in a device consisting of a branching planar waveguide integrated with a microliquid handling system [48] (Fig. 4.29):

A waveguide branches into a sensing arm and a reference arm. In the sensing arm, a region of the waveguide-cladding layer is removed, so the evanescent field is exposed to the sample. Evanescent-wave sensing can be performed in several ways. The simplest approach is based on absorption measurement in which a wavelength is chosen that corresponds to an absorption peak of the sample. This results

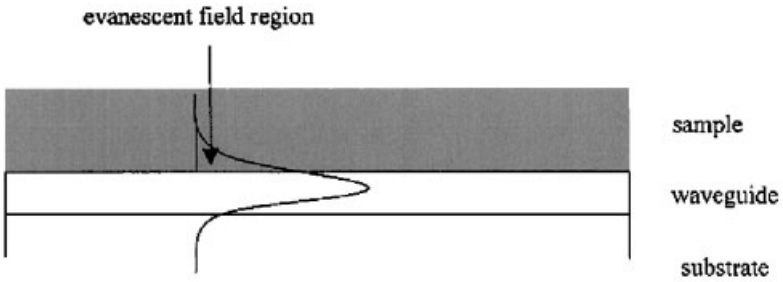


Fig. 4.28 Principle of evanescent-wave sensing. The evanescent field that extends outside the waveguide core is used for probing the sample [48].

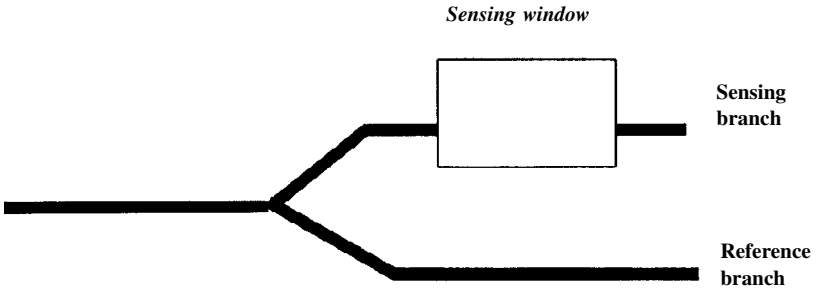


Fig. 4.29 Configuration of the optical system in a device based on evanescent-wave sensing. The sensing window is 1000 μm long.

in attenuation of the optical power in the sensing arm compared to the reference arm, which can be described by the Lambert–Beer law (Eq. 4.8 and Eq. 4.9).

Evanescent-wave sensing of nonabsorbing samples can also be done by measuring the change in refractive index, which also depends on the solute concentration. This approach relies on measurements of the interference pattern between the sensing branch and the reference branch. It can be orders of magnitudes more sensitive than traditional absorbance measurements, but is unfortunately also difficult to control, because other factors such as the temperature influence the refractive index of the sample. The optical readout is also more complicated.

4.4.4 Fluorescence Detection

Fluorescence spectroscopy involves excitation of an electron to a higher-energy state due to absorption of a photon (see section 4.4.2), followed by relaxation of the electron to the lower-energy states (Fig. 4.30) resulting in emission of photons in addition to dissipation of thermal energy to the surroundings. When a laser is used for excitation, the detection method is called laser-induced fluorescence (LIF).

The emission spectrum is always shifted towards longer wavelengths than the absorption spectrum (Fig. 4.31), because energy is conserved throughout the excitation and relaxation processes. Because a fraction of the excitation energy is typically dissipated as thermal energy (nonradiative relaxation), less energy is left for emission of photons (radiative relaxation); thus, the emission spectrum is located at lower energies than the absorption spectrum. This corresponds to a longer wavelength, because energy and wavelength are inversely proportional, as seen in the well-known formula:

$$E = h\nu = \frac{hc}{\lambda} \tag{4.10}$$

For calculations it is easier to remember the formula as

$$E[eV] = \frac{1.24}{\lambda[\mu m]} \tag{4.11}$$

where energy is in units of electron volt and wavelength is in micrometers.

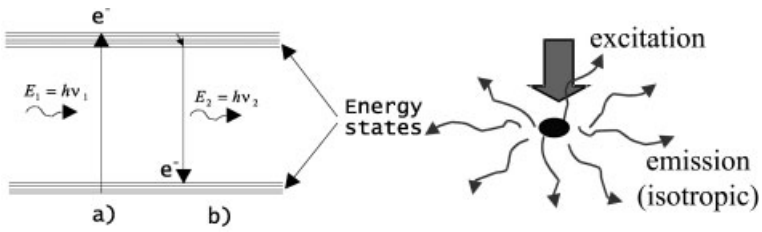


Fig. 4.30 Left: Energy diagram showing the concept of fluorescence. A photon with energy E_1 is absorbed, and a photon with lower energy E_2 is subsequently emitted. Right: Schematic drawing showing the isotropic nature of the emitted light.

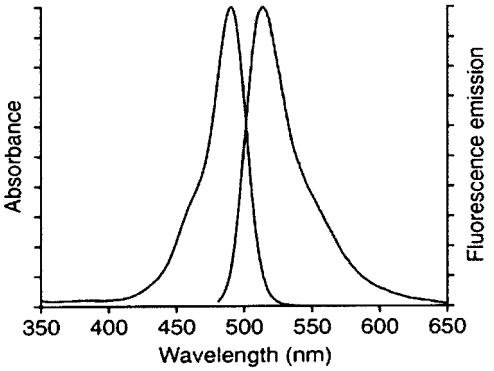


Fig. 4.31 Absorption (left) and emission (right) spectra of fluorescein, a widely used fluorophore.

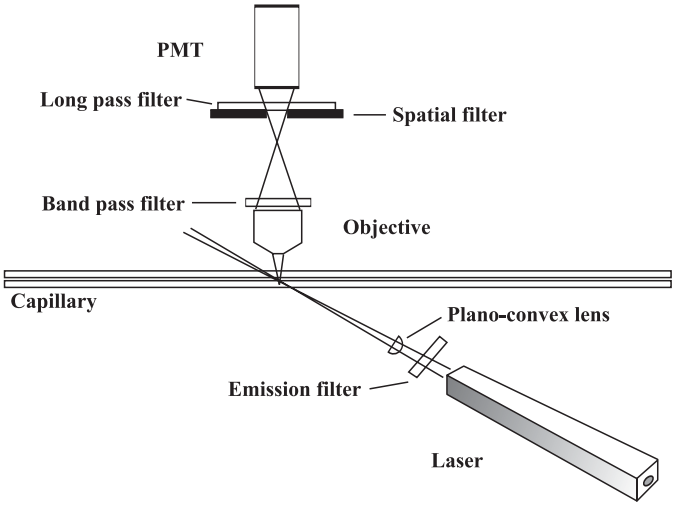


Fig. 4.32 Schematic drawing of a typical optical setup for fluorescence measurements [49].

Fluorescence measurement is widely used because very low limits of detection can be achieved – it is even possible to measure single molecules. A disadvantage compared to other techniques such as absorption measurement is that the optical setup is relatively complex, because the emitted light has to be separated from the excitation light, since collection of excitation light results in an increase in the background signal and hence an increase in the noise of the measurement. A schematic drawing of a typical setup is shown in Fig. 4.32.

A laser is focused onto a section of a fluidic channel from an angle of about 45° to avoid shining its light directly into the photodetector. The fluorescence is collected by an objective lens before which filters, such as bandpass and long wave pass filters, are inserted in the light path to suppress collection of excitation light. A spatial filter is also typically used to ensure that only light from a well-de-

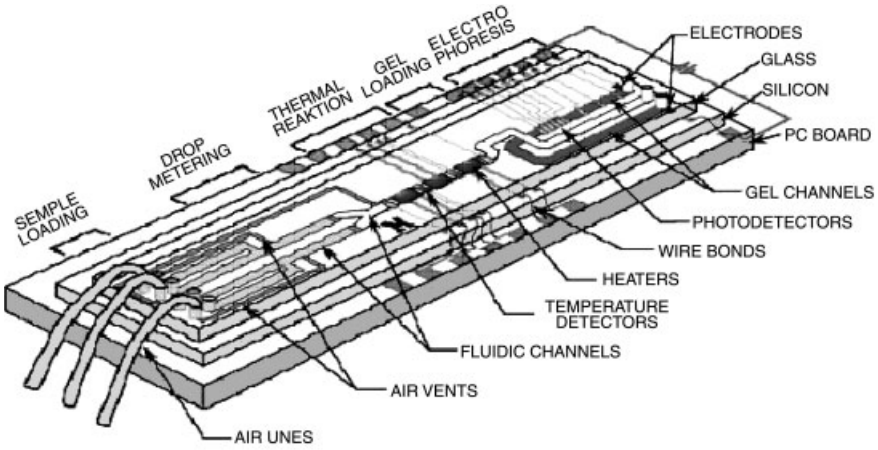


Fig. 4.33 Microfabricated device for DNA analysis with an integrated fluorescence detector [51].

fined region of the channel is collected. Hence, many different optical components are used. The setup is also bulky and very sensitive to correct alignment of the excitation light and the collection optics.

One of the most impressive devices in terms of integration was developed for fluorescence measurements in DNA analysis [50] (Fig. 4.33). This device was fabricated from a silicon substrate with integrated photodiodes for optical detection. A thin film interference filter was deposited on top of the diode to suppress measurement of excitation light. Optical excitation was applied from the top of the device with an external blue-light-emitting diode (LED). The fluidic channel network consisted of a sample loading compartment, a PCR chamber for DNA amplification, and channels for separation of the DNA molecules by capillary electrophoresis.

4.5 Electrochemical Sensors

Electrochemical sensors can often be applied if the sample is for instance turbid. Another advantage is that electrochemical sensors are relatively easy to fabricate because they require only two electrodes that normally consist of noble metals such as platinum, silver, or gold. Other electrode materials used are vitreous/glassy carbon or graphite or, more advanced because optically transparent, indium tin oxide (ITO). The measurement principles most often applied are potentiometry and amperometry. As the name indicates, a potential or a current is measured while keeping the current flow or the potential difference between two electrodes constant.

One important aspect of each electrochemical measurement is to chose and fabricate the right reference electrode. The purpose of a reference electrode is to have a fixed potential to relate the signal to, because only potential differences can be measured. Conventionally, all redox potentials were related to the normal hydrogen electrode (NHE), an electrode that consists of a platinum mesh, dipped into 1 M sulfuric acid, purged with 1.013 bar hydrogen at a temperature of 25 °C. The potential of the NHE was set to 0 V.

Because this type of electrode is not very convenient, other stable reference electrodes were invented. Common, so-called second-generation, reference electrodes consist of a metal such as mercury or silver and a corresponding, barely-soluble salt such as Hg_2Cl_2 (calomel) or AgCl. The purpose is to create a saturated solution of the ions, resulting in a stable reference potential. However, this reference potential is still temperature dependent, for instance, the potential of a Ag/AgCl system with saturated KCl solution is 220.5 mV at 0 °C and 159.8 mV at 60 °C [51].

Because calomel electrodes are somewhat environmentally problematic, Ag/AgCl is used much more today. It is actually unimportant which system is used – both types have the same quality.

Several researchers have tried recently to fabricate stable and accurate miniaturized reference electrodes; however, most were unstable or inaccurate. The major flaw lies in the 'consumption' of either the barely-soluble salt or the electrode material. A thin-film electrode consisting of, for instance, only 1000 nm Ag with a surface area of less than a square millimeter completely dissolves within a couple of weeks according to the reaction:

$$Ag \rightarrow Ag^+ + e^- \tag{4.12}$$

and a small volume of solid AgCl dissolves very quickly if washed with deionized water for even a short period. This of course affects the stability of that reference electrode, resulting in a drifting and most often unstable signal.

However, microscale Ag/AgCl systems intended for short-term use (<2 weeks working condition) can be built by relatively simple techniques: an Ag electrode (e.g., thin film Ag deposited on a silicon chip) is dipped into a chloride solution and covered electrochemically at potentials of ~150 mV against a commercial Ag/AgCl reference electrode, or AgCl can be deposited out of an iron(III) chloride solution according to the following equation:

$$FeCl_3 + Ag \rightarrow AgCl + FeCl_2 \tag{4.13}$$

Slight modifications of the electrode setup offer a variety of analytical possibilities – the so-called ion-selective electrodes (ISE). As an example, the fabrication of a potassium-selective electrode is given in section 10.1.

To miniaturize potentiometric sensors, several attempts have been made. One of the most elegant techniques for making miniaturized electrochemical systems is based on the ion-selective field effect transistor (ISFET). Bergfeld et al. [52] removed the metal oxide of a metal oxide field effect transistor (MOSFET) and re-

placed it with an ion-selective membrane mix. In general, the functional principle of an ISFET is similar to that of a triode. Triodes were used around the 1960s in radios and later in televisions.

In a *p*-type doped silicon wafer, two regions on the top are doped with electron donators such as nitrogen or phosphorus. Between those two regions, called source (S) and drain (D) of the FET a thin insulating layer is deposited. Commonly used materials are silicon oxide or silicon nitride. If the FET is connected as in Fig. 4.34, a thin electron channel between source and drain is developed right at the interface between the insulator and the bulk *p*-doped silicon.

Depending on the potential difference between the gate and the bulk silicon, this electron channel becomes more or less deep, thereby influencing its conductance. This is known as the field effect. Because the conductance change itself influences current flow between source and drain, this current flow is directly proportional to the gate potential. In principal, an FET with a gate made of silicon nitride or tantalum oxide is sensitive towards protons and can therefore be considered an ISFET that senses the pH value. If the gate is now covered with an ion-selective membrane, other ions such as ammonium, potassium, sodium, nitrate, and many others can be measured.

The first commercial ISFETS were pH-sensitive due to a thin layer of tantalum oxide or silicon nitride – both materials are frequently used as masking materials in silicon processing and therefore are well known to processing engineers. An additional advantage is the adhesion of these inorganic layers: they stick to the gate surface much better than the above-mentioned liquid membranes.

After the first pH-ISFETS entered the market, several ISFET-type sensors were developed and commercialized. The 'classic' pH-ISFET is sold as an 'unbreakable' alternative to the glass electrode. However, because it is machined of crystalline silicon, this material also can break, and additionally, measurement data obtained from ISFETs tend to drift. The main application of pH-ISFETs is in the food industry: the sensor can for example be inserted directly into an apple or a tomato – an advantage compared to the thin glass membrane of a glass electrode, which would break immediately. The advantage of being small with respect to the glass electrode has, surprisingly enough, not resulted in many commercial products yet. The reason for this is most likely one of the still unsolved problems in elec-

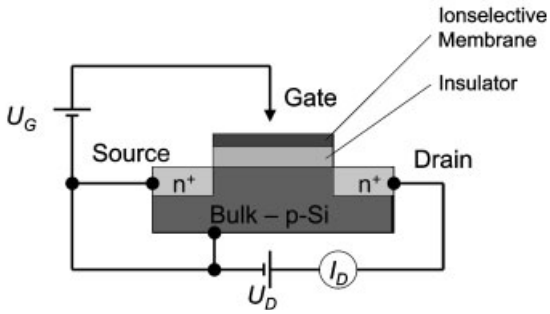


Fig. 4.34 Schematic drawing of ion-selective field effect transistor (ISFET) [53].

trochemistry: constructing a functional miniaturized reference electrode. One approach to circumvent the necessity of a reference electrode so as to take advantage of miniaturized sensors is to use differential techniques for the measurement. Instead of a reference electrode, a second so-called reference-FET or ReFET can be used. However, it has to be demonstrated that both FETs (ISFET and ReFET) have the same slope and drift in signal.

4.6 References

1 Kovacs, G.T.A. *Micromachined transducers sourcebook*; McGraw-Hill: New York, 1998.
2 Woias, P. *Proceedings of SPIE*, San Francisco 2001; SPIE; 39–52.
3 Shoji, S. In *Microsystem Technology in Chemistry and Life Science*; Becker, H., Manz, A., Eds., 1998; Vol. 194, pp 164–188.
4 Madou, M. *Fundamentals of microfabrication*, first ed.; CRC Press: Boca Raton, Florida, 1997.
5 Yu, Q.; Bauer, J.M.; Moore, J.S.; Beebe, D. J. *Applied Physics Letters* **2001**, *78*, 2589–2591.
6 Beebe, D. J.; Moore, J.S.; Bauer, J.M.; Yu, Q.; Liu, R.H.; Devadoss, C.; Jo, B.H. *Nature* **2000**, *404*, 588–590.
7 Krog, J.P.; Dirac, H.; Fabius, B. et al. Proceedings of the μTAS 2000 Symposium, Enschede, Netherlands, May 14–18, 2000; Kluwer Academic Publishers; 419–422.
8 Duffy, D.C.; Gillis, H.L.; Lin, J.; Sheppard, N.F., Jr.; Kellogg, G.J. *Analytical Chemistry* **1999**, *71*, 4669–4678.
9 Ekstrand, G.; Holmquist, C.; Örlefors, A.E.; Hellman, B.; Larsson, A.; Andersson, P. *Micro Total Analysis Systems 2000: Proceedings of the μTAS 2000 Symposium*, Enschede, The Netherlands, May 14–18, 2000; Kluwer Academic Publishers; Dordrecht; 311–314.
10 Thomas, N.; Ocklind, A.; Blikstad, I.; Griffiths, S.; Kenrick, M.; Derand, H.; Ekstrand, G.; Ellström, C.; Larsson, A.; Andersson, P. *Proceedings of the μTAS 2000 Symposium*, Enschede, The Netherlands, May 14–18 2000; Kluwer Academic Publishers; Dordrecht; 249–252.
11 Adamson, A.W. *Physical Chemistry of Surfaces*; John Wiley and Sons: New York, 1990.
12 Schwer, C.; Kenndler, E. *Analytical Chemistry* **1991**, *63*, 1801–1807.
13 Li, S.F.Y. *Capillary Electrophoresis*; Elsevier: Amsterdam, 1993.
14 Paul, P.H.; Garguilo, M.G.; Rakestraw, D.J. *Analytical Chemistry* **1998**, *70*, 2459–2467.
15 Guo, Y.; Imahori, G.A.; Colon, L.A. *Journal of Chromatography A* **1996**, *744*, 17–29.
16 Barker, S.L.R.; Tarlov, M. J.; Canavan, H.; Hickman, J.J.; Locascio, L.E. *Analytical Chemistry* **2000**, *72*, 4899–4903.
17 Barker, S.L.R.; Ross, D.; Tarlov, M.J.; Gaitan, M.; Locascio, L.E. *Analytical Chemistry* **2000**, *72*, 5925–5929.
18 Schasfoort, R.B.M.; Schlautmann, S.; Hendrikse, J.; van den Berg, A. *Science* **1999**, *286*, 942–945.
19 Ramsey, R.S.; Ramsey, J.M. *Analytical Chemistry* **1997**, *69*, 1174–1178.
20 Lazar, I.M.; Ramsey, R.S.; Jacobson, S.C.; Foote, R.S.; Ramsey, J.M. *Journal of Chromatography A* **2000**, *892*, 195–201.
21 Culbertson, C.T.; Ramsey, R.S.; Ramsey, J.M. *Analytical Chemistry* **2000**, *72*, 2285–2291.
22 McKnight, T.E.; Culbertson, C.T.; Jacobson, S.C.; Ramsey, J.M. *Analytical Chemistry* **2001**, *73*, 4045–4049.
23 Alarie, J.P.; Jacobson, S.C.; Broyles, B.S.; McKnight, T.E.; Culbertson, C.T.; Ramsey, J.M., Monterey, California, USA 2001; Kluwer Academic Publishers; 131–132.
24 Paul, P.H.; Arnold, D.W.; Rakestraw, D. J., Banff, Canada 1998; Kluwer Academic Publishers; 49–52.

25 Paul, P.H.; Arnold, D.W.; Neyer, D.W.; Smith, K.B. In *μTAS 2000*; van den Berg, A., Olthuis, W., Bergveld, P., Eds.; Kluwer Academic Publishers: Monterey, California, 2000, pp 583–590.

26 Zeng, S.; Chen, C.-H.; Mikkelsen, J.C.; Santiago, J.G. *Sensors and Actuators B* **2001**, *79*, 107–114.

27 Stroock, A.D.; Dertinger, S.K.W.; Ajdari, A.; Mezic, I.; Stone, H.A.; Whitesides, G.M. *Science* **2002**, *295*, 647–651.

28 Johnson, T. J.; Ross, D.; Locascio, L.E. *Analytical Chemistry* **2002**, *74*, 45–51.

29 Oddy, M.H.; Santiago, J.G.; Mikkelsen, J.C. *Analytical Chemistry* **2001**, *73*, 5822–5832.

30 Jacobson, S.C.; Koutny, L.B.; Hergenröder, R.; Moore Jr., A.W.; Ramsey, J.M. *Analytical Chemistry* **1994**, *66*, 3472–3476.

31 Jacobson, S.C.; Hergenröder, R.; Moore Jr., A.W.; Ramsey, J.M. *Analytical Chemistry* **1994**, *66*, 4127–4132.

32 Alarie, J.P.; Jacobson, S.C.; Culbertson, C.T.; Ramsey, J.M. *Electrophoresis* **2000**, *21*, 100–106.

33 Zhang, C.X.; Manz, A. *Analytical Chemistry* **2001**, *73*, 2656–2662.

34 Laurell, T.; Wallman, L.; Nilsson, J. *Journal of Micromechanics & Microengineering* **1999**, *9*, 369–376.

35 Petersson, M.; Nilsson, J.; Wallman, L.; Laurell, T.; Johansson, J.; Nilsson, S. *Journal of Chromatography B*. **1998**, *714*, 39–46.

36 U.D. Larsen, *Micro Liquid Handling*, PhD Thesis, Mikroelektronik Centret, Kgs. Lyngby, Denmark, July 2000.

37 C. J. Yeager and S.S. Courts, *IEEE Sensors Journal*, **2001**, 1(4), 352–560.

38 A.F.P. van Putten, *Electronic Measurement Systems*, 2nd edit, Institute of Physics Publishing, Bristol, **1996**.

39 Z. Zhan, C. Dafu, Y. Zhongyao, W. Li, *Biochip for PCR amplification in silicon*, **2000**, 1st Annual International IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine & Biology, Lyon, October 12–14.

40 P. Hauptmann, *Sensoren: Prinzipien und Anwendungen*, Carl Hanser, Munich, **1990**.

41 R. Frank, *Understanding Smart Sensors*, Artech House, Boston, **1996**.

42 Y. Sato, G. Irisawa, M. Ishizuka, K. Hishida and M. Maeda, *Measurement Science and Technology*, **2003**, 14, 114–121.

43 S.W. Allison and G.T. Gillies, *Review of Scientific Instruments*, **1997**, 68 (7), 2615–2650.

44 A.M. Chaudhari, T.M. Woudenberg, M. Albin and K.E. Goodson, *Journal of Microelectromechanical Systems*, **1998, 7**(4), 345–355.

45 G.C.M. Meijer and A.W. van Herwaarden, ed, *Thermal Sensors*, Institute of Physics Publishing, Bristol, **1994**.

46 R. Marie, J. Thaysen, C.B.V. Christensen and A. Boisen, *Microelectronic Engineering, 1,* **2003**, 67–68, 893–898.

47 K.B. Mogensen et al. *Electrophoresis*, 22 (**2001**), p 3930.

48 G. Pandraud et al. *Sens. Actuat. A*, 85 (**2000**), p 158.

49 S.A. Shippy et al., *Analytica Chemica Acta* 307 (**1995**).

50 Burns et al., *Science*, **1998**.

51 H.R. Christen, *Grundlagen der allgemeinen und anorganischen Chemie*, Verlag Sauerlaender Ag Aarau, 9. Auflage **1988**.

52 P. Bergveld, *Ion sensitive field effect transistor*, International Conference on Biomedical Transducers, **1975** pp 299–304.

53]. K. Cammann, *Instrumentelle Analytische Chemie*, **2000**, Spektrum Akademischer Verlag, Heidelberg, Berlin.