Professor Gary Fedder, Integrated Micro Systems (18-819)
Carnegie Mellon University, Fall 1996

A Halitosis Sensor

for the quantitative assessment of oral malodor in humans

by Marc Bohlen

Abstract: A novel method of measuring the main constituent of bad breath, hydrogen sulfide, is described. The proposed design uses a flexural plate wave device to transduce the concentration of hydrogen sulfide. The calculated sensitivity to hydrogen sulfide is -175 cm^2/gr. Cross sensitivity to other gases, mainly hydrogen, is avoided by temperature control through integrated resistive heaters and thermocouples. The device can be implemented in a DIMES similar production process.

Introduction:

Halitosis is common among adults. Humans are quite sensitive to halitosis in others but unable to assess it in their own breath. In an appearance conscious society this often leads to uncomfortable situations. In clinical dentistry halitosis is measured to monitor dental hygiene.
The main chemical constituents of oral malodor are volatile sulfides. Hydrogen sulfide (H2S) and methyl mercaptan (CH3SH) are the most pronounced[1]. Traditionally, the hydrogen sulfide content of breath is measured with gas chromatography and the concentration thereof related to the degree of malodor. Even though this approach has been questioned it is the bad breath measurement of choice. Comparisons with human malodor experts show that the threshold for bad breath lies around 0.12ppm[2].
Instruments for gas chromatography are bulky and expensive[3]. With MEMS one now has the opportunity to build miniature and inexpensive IC compatible devices that facilitate routine measurements.

Physical principles of the proposed sensor design:

A practical transducer for monitoring gas concentrations is the mass loading effect in flexural plate wave (FPW) devices. A material that is sensitive to the sought measurand is coated on the plate and an absorption process occurs when the gas contacts the plate. A mechanical wave moving through a plate that is in contact with the gas of choice will have a lower velocity than in an isolated plate. This velocity difference is linearly dependent on the gas concentration and can be used to monitor the concentration changes[4].
In FPW devices the lowest order antisymmetric mode (Ao) that involves flexure allows the most sensitive measurements[5]. The resulting Lamb waves require a plate thickness that is much smaller than the wavelength of operation. Lower propagation velocities result in longer delay times. As the velocity is dependent on the overall thickness of the plate, thin and large plates are preferred for Lamb wave devices. ECE (electrochemical etching) bulk back etching is a proven method to achieve very thin membranes with small variations in thickness. Mathematically, the wave velocity is given by:

$$vp(\lambda) = \left[\left(T + (2\pi/\lambda)^2 Ed^3\right)/\left(M \times 12\left(1-\upsilon^2\right)\right)\right]^{1/2} \quad [6] \quad [1]$$

where T is the residual stress, E is Young's modulus, λ the wavelength of operation, d the overall plate thickness, M the plate mass and υ the Poisson's ratio.
The perturbation of the wave velocity is:

$$\Delta v/vo = (1/vo)\left[(\delta v/\delta m)\Delta m + (\delta v/\delta T)\Delta T + (\delta v/\delta\sigma)\Delta\sigma + (\delta v/\delta\varepsilon)\Delta\varepsilon + ...\right]^{7} \qquad [2]$$

where vo is the original, undisturbed velocity of the Lamb wave, T the ambient temperature, σ the conductivity and ε the dielectric coefficient.
As equation [2] shows, there are many factors related to the change of velocity. It is the designer's task to devise a device that uses one of the parameters as the dominant effect. In the case of coated FPW device, the dominant parameter is mass. A change in mass induces a change in velocity which generates a different delay time or a change in the output frequency:

$$\Delta m \rightarrow \Delta v \rightarrow \Delta t \rightarrow \Delta f \qquad [3]$$

Tungsten trioxide (WO3) has been found to show a reliable sensitivity to H2S[8]. The nature of the interaction remains somewhat unclear but a redox reaction is plausible[9]. WO3 shows, in contrast to lead acetate, long-term stability and can be sputtered into thin, dense films. WO3 is therefore the H2S sensitive material of choice.

Aluminum interdigital transducers (IDTs) are arranged such that energy in the form of a time varying voltage applied to the IDTs is converted via a piezoelectric material into mechanical energy, the Lamb wave described above. Zinc oxide (ZnO) is a common piezoelectric material and can be used for this purpose since it, too, can be applied in very thin, dense films[10].
The critical factor for effective energy conversion is the coupling factor:

$$\kappa^2 = e^2/\varepsilon c \qquad [11] \qquad [4]$$

where e is the piezoelectric coefficient, ε the dielectric coefficient and c the stiffness coefficient. The coupling factor is often evaluated experimentally as:

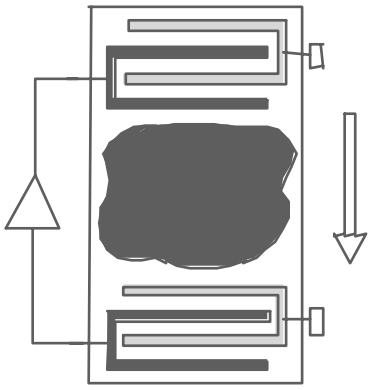
$$k_{eff}/2 = \Delta v/vo \qquad [12] \qquad [5]$$

ZnO does not have an exceptionally high coupling factor. However, if a highly oriented film can be grown in the c-axis perpendicular to the substrate, then acceptable coupling factors can be achieved. For thin films between 0.2 and 0.7 um on devices operating around 10 Mhz coupling factors of 0.02 to 0.04 have been reported[13], provided the ratio between ZnO and the underlying silicon substrate is optimized.

Delay line oscillators for gas sensors:

As the process by which electrical energy is converted into mechanical energy is accompanied by high losses, one must add an amplifier element to the FPW device to ensure that the output of the IDTs can be used as a new input.

A(f) B(f)



Due to the above mentioned losses the amplifier must supply a loop gain A(f) > 1. For continuous oscillation we must have:

$$|A(f)||B(f)|=1 \qquad [6]$$

The loop phase is equal to:

$$2\pi n = \phi_{delayline} + \phi_{amplifier} + \phi_{IDT1} + \phi_{IDT2} \qquad [7]$$

[14]

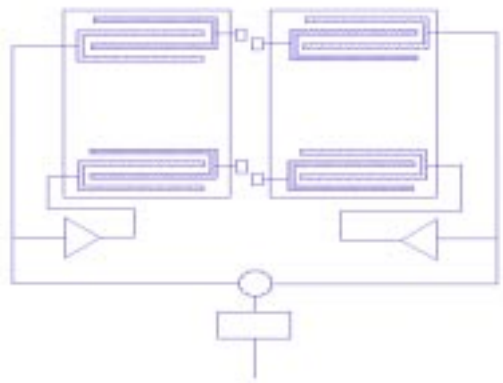
Since the phase of the delay line is much larger than the phase of the rest of the elements:

$$\phi_{delayline} = 2\pi fL / v \qquad [8]$$

and the operating frequency is:

$$f_{op} = \frac{2\pi n}{2\pi(L/v)} = v_{phase} / L_{delayline} \qquad \text{when n=1} \qquad [9]$$

In order to minimize the effects of pressure variation and temperature drift the FPW device will be operated in a dual delay line oscillator configuration:



The device on the left is the reference plate and is left uncoated. The device on the right is coated with the H2S sensitive WO3 on the bottom of the ECE etched plate. At the amplifier outputs, the respective frequencies are picked off, mixed (subtracted) and cleaned in the last stage, a low pass filter. If one assumes that the devices are fabricated identically and that the only difference arises from the coated/uncoated plate, then this configuration effectively cancels common mode effects.

Design constraints:

In order to easily measure the delay time difference between the coated and uncoated FPWs one would like to have long delay times, long delay lines and thus large devices. However, the sensitivity and the rise time due to thermal properties are improved with a small device. As the interaction between H2S and WO3 is best between 130 and 200 C, the device must be thermally controlled. This can be done by integrating resistive heaters and thermocouples. In order to minimize the energy required to heat the device and to optimize the thermal rise time the FPWs will be thermally isolated in a SI3N4 diaphragm. As mentioned above the aluminum contacts deposited on top of the ZnO piezoelectric layer are interdigitized fingers. The Lamb waves interfere constructively if the distance between adjacent fingers is equal to one half of the wavelength of operation[15]. Maximum positive interference occurs at the resonant frequency. At all other frequencies the emitted waves partially cancel each other, reducing the vibration. The bandwidth of the output is dependent mainly on the number of transducer fingers. Increasing the number of fingers decreases the bandwidth. In order to minimize the device capacitance the number of fingers should be small. However, in order to minimize the conversion losses the number of fingers should be large. Diffraction losses can be held small if the aperture (the length of the fingers) is large. A length of 20 λ is sufficient for this purpose.
As the surface waves are emitted in opposite directions one has an inherent minimal loss of 3 dB per IDT (6dB for both IDTs). A number of iterative calculations that balance these partially conflicting requirements resulted in the values listed below.

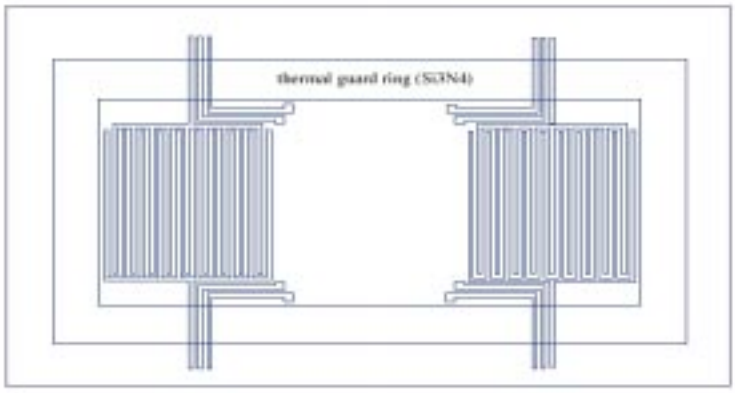
FPW design:

Calculated key parameters for design of the Lamb wave device:

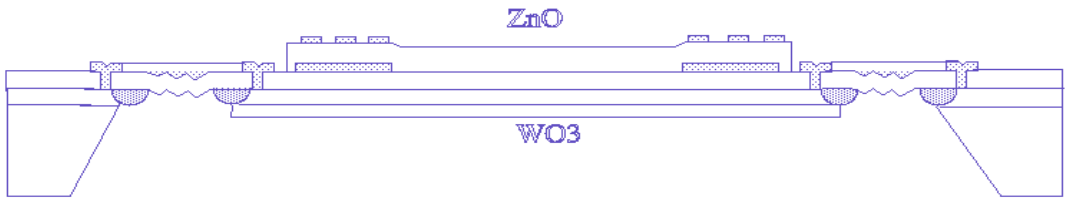
overall dimensions	2560 x 5960 um
thickness of coated plate	7.5 um
mass of coated plate	786 nkg
wavelength of operation	80 um
# of transducer fingers	20 pairs
spacing between fingers	40 um
aperture (length of fingers)	2000 um
length of delay path (between IDTs)	3000 um
(corrugated) Si3N4	500 um on each side
phase velocity	960 m/s
operating frequency	12 Mhz
delay time (between IDTs)	3.13 us
sensitivity	-175 cm^2/gr

The layout and a side view are included below.

The calculated sensitivity is about twice as high as in comparable SAW devices operating at 100 Mhz, but about a third of the value attainable in similar FPW devices[16]. This is a result of the increased plate thickness due to the integrated resistors and the WO3 film.



The layout shows the device version without corrugation .

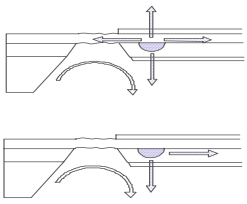


material	thickness [um]
wafer	500

n-epi	4
Si3N4	2
AL for ground plate	0.25
ZnO	0.5
AL for IDTs	0.25
WO3	0.5

heater and thermocouple design:

The heater is realized with an embedded p+ (boron) doped polysilicon element[17]. The resistor is meandered to increase the interface area with the surrounding material. The thermocouple is made of aluminum and p+ doped polysilicon. In order to estimate the thermal properties of the device a number of assumptions were made. As the device will be in contact with human breath a small amount of heat will be lost to (non forced) convection. Furthermore, the 500 um silicon nitride ring will thermally isolate the FPW preventing heat flow into the surrounding thick rim[18],[19]. The values below are based on a temperature increase of 130 C[20].

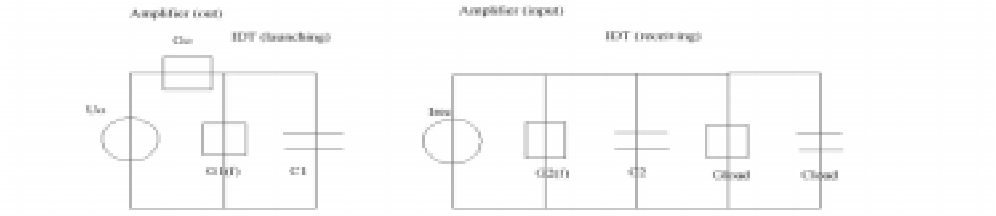


	resistive heater	thermocouple
doping	boron	boron
length	6000 um (meander)	3000 um
width	30 um	15 um
	resistive heater	thermocouple
depth	3 to 4 um	3 to 4 um
thermal resistance	2400 K/W	
required power	60 mW	
measured output		70 mV*
time constant	0.5 sec	
rise time (95%)	1.5 sec	

*assuming a Seebeck coefficient of 500 uV/K for p+ polysilicon. (Semiconductor Sensors, S.M.Sze, p. 354.)

equivalent circuit and amplifier requirements:

The reason for modeling the IDTs with an equivalent circuit is to be able to make a wise choice in the amplifier design for the oscillator. In the simplified model below there is no inductance term.



If G0 is much greater than |G1 + jwC1(w)| and C2(w) is much greater than CLoad:

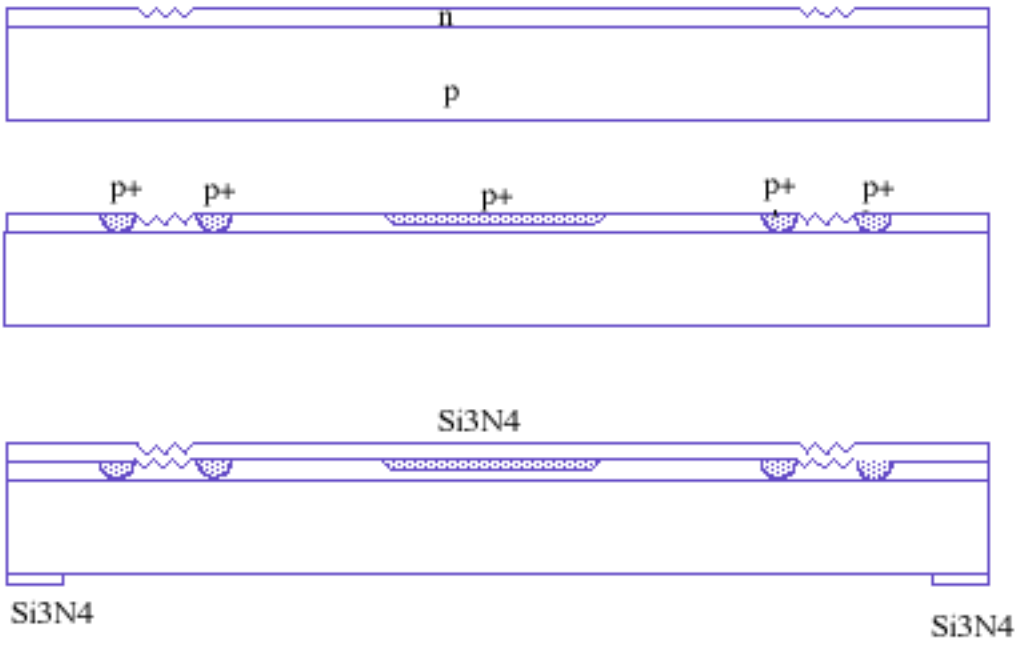
$$\frac{U_L^2}{U_o^2} = \frac{G_1(w)G_2(w)}{\left(G_L + G_2(w) + jwC_2\right)^2} \qquad [10]$$

The inductance of the load, the amplifier, is usually the smallest of the terms in the denominator and can be neglected. A voltage sensing amplifier design will have a high common mode rejection ratio up to high frequencies and is the recommended approach for oscillators of this kind. As mentioned above, the inherent losses due to bi-directional propagation are 6dB. However, the insertion losses due to mechanical coupling are expected to be large. Typical values are 10 to 15 dB each for the launching and the receiving IDT. The total losses are on the order of 30 to 40 dB. The amplifier must be able to handle this dynamic range. Since the gain requirements to the amplifier vary, an automatic gain control circuit is often included[22]. This addition ensures a constant overall gain of unity under all operating conditions.
A similar control loop is required for the heater and heat sensing element such that the operating temperature remains in the optimum range throughout the sensing procedure.

production process:

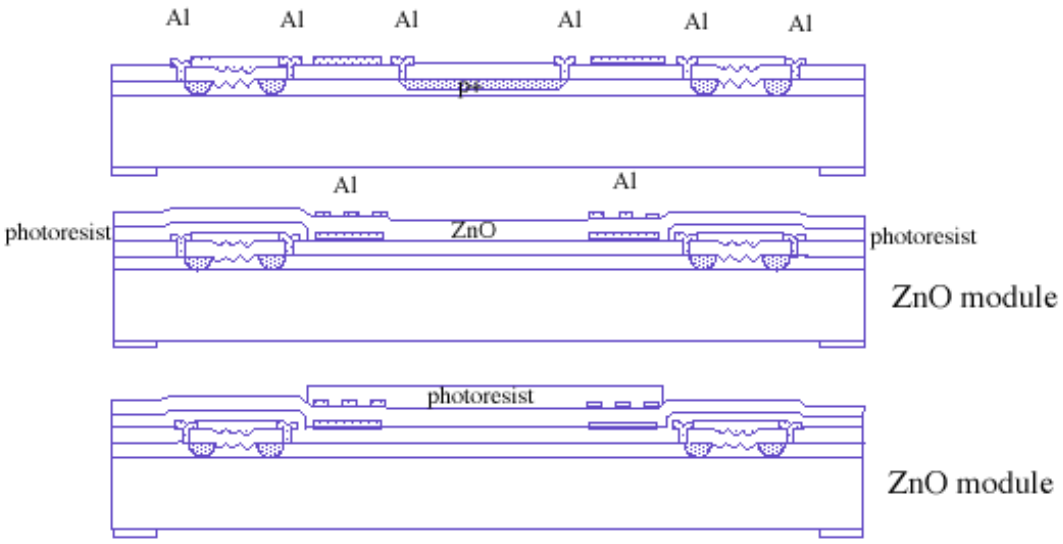
In order to minimize the production cost it is important that the fabrication process remain as compatible as possible with standard IC processes. As ZnO and WO3 do not

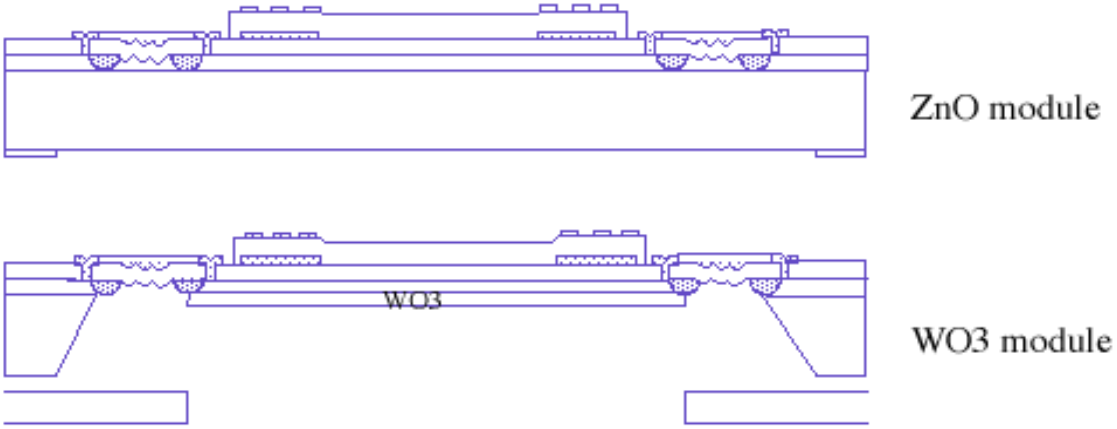
meet this requirement, separate process modules should be created to apply these materials. The Technical University of Delft has developed a fabrication scheme that takes many of these considerations into account[23]. Their DIMES process will therefore be the basis of the production process described here. A sequence of cross sections describing the process is included below. (The cross sections 5 to 8 do not include the heater elements as not to clutter the diagrams.)
The IC process begins with a double side polished p type <100> silicon wafer of 500 um on which an n-epilayer of 4 um is deposited. For the corrugated version of the thermal guard ring, this layer is first patterned and etched in EDP @ 115 C (etch rate 1 um/min). Then p+ boron is diffused into the n-layer to create the heater resistor, the thermocouple and the guard ring . The latter element is used during the ECE process to electrically isolate a 500 um wide thermal buffer of silicon nitride around the FPW. It is therefore important that the depth of the guard ring element meet the n-epi layer thickness of 4 um. Thereafter a 2 um thick layer of Si3N4 is deposited by mixed frequency plasma enhanced CVD. By varying the deposition frequency a nearly stress free layer can be created.



Then vias through the SI3N4 to the doped poly are formed. 0.25 um of aluminum are deposited and patterned for the ECE connect, the ground plates of the IDTs and the connectors for the heaters and thermocouples.
A sacrificial layer of photoresist is deposited and patterned over the area that will not be coated with ZnO. A 0.5 um thick zinc oxide layer is then D.C. sputtered onto the plate @375 C at a rate of 100nm/min. ZnO will exhibit considerable residual stress if the deposition parameters are not optimized. Since the temperature coefficient of zinc oxide (4x10^-6/K) is larger than that of silicon (2.3x10^-6/K) a residual tensile stress due to

cooling after deposition will be present. This tensile stress can be compensated by creating a compression stress component. If the film is sputtered at a pressure of 6mTorr and 375 C, zero residual stress interface can be created[24].
Thereafter the IDTs are created by vacuum evaporating 0.25 um aluminum onto the ZnO. The IDTs are formed by etching in KOH and potassium iron cyanide (K3Fe(CN)6) @ 20 C (etch rate 0.2 um/min). The masking material is photoresist. The ZnO layer is then wet etched with a phosphoric acid mix: H3PO4, HOAc (Hydrogen Acid) and H2O (1:100:100) @ 20 C (etch rate 0.2 um/min). The masking material is photoresist and is removed, as above, with acetone.
At this moment the wafer is ready for the electrochemical backside etch. The wafer is exposed to KOH @ 85 C to form the diaphragm. This etch process is not timed and is terminated when the etchant has removed the p-layer completely and disrupted the diode that enabled the ECE etch. [25] As a ring of 500 um Si3N4 is electrically isolated by the p+ doped guard lines the wafer will etch down to the silicon nitride in this region. Since silicon nitride is a very good thermal isolator (k = 0.3W/mK) the device will be thermally isolated from the surrounding wafer material.
The WO3 module is the last step. The backside is shadow masked as shown on the next page and 0.5 um of Au-doped WO3 is r.f. sputtered onto the plate. In order to increase the sensitivity of the WO3 film it is thermally activated @ 250 C, 400 C and then cooled.





Assessment of device performance:

- sensitivity:

J.F.Vetelino designed a SAW H2S sensor in 1987 with a sensitivity of 10 ppb[26]. Since the proposed FPW device operating at 12 Mhz has a sensitivity that is about twice as large as that of Vetelino's 60 Mhz SAW it should meet and exceed the required measurement limit of 120 ppb to detect the critical concentration of H2S. Vetelino's device measured frequency differences in the range 1 to 10 kHz . As it is difficult to evaluate the amount of H2S that will interact with the WO3 layer, a quantitative assessment of the change in mixed frequency output is not possible. Using the same argument as above one can be sure that the frequency change will at least as large as the one measured by Vetelino.

- bad breath:

Measuring H2S is not really assessing halitosis. As bad breath really is a mixture of a number of gases one should make an array of such FPW devices, each sensitive to a particular component of halitosis process the outputs to a characteristic "signature" of bad breath.

- cross sensitivity to other gases and repeated measurements:

Since the WO3 film was thermally activated after its deposition the only other gas it is sensitive to is hydrogen (H2)[27]. As H2 shows a peak sensitivity at about 230 C and H2S shows peak sensitivity below 200 C it is easy to discriminate between the two gases. The proposed device should therefore be operated at 130 to 200 C. This corresponds to a thermocouple output of 70 to 87 mV.

The WO3 H2S interface can be “degassed” by shortly heating the device to temperatures above 200 C. The process is reversible and the device thus reusable.

- thermal rise time:

Due to the thermal isolation provided by the silicon nitride ring the device needs only about 60 mW to heat up to the required operating temperature. The rise time is about 1.5 sec. However, the peak sensitivity between WO3 and H2S is only reached after about 2 minutes[28]. This is the limiting factor in the speed of operation.

- mechanical stability:

As described above the deposition process can be manipulated to create close to zero stress plates at room temperature. The critical buckling stress for diaphragms is given by:

$$S_c = 5.33\frac{\pi^2}{a^2 h}\frac{Eh^3}{12(1-\nu^2)}$$ [29] [11]

Where E is Young’s modulus, h is the thickness of the diaphragm, a^2 is the area and υ is Poisson’s ratio. Numerically this equates to about 2.5 MPa for the proposed design, obviously much larger than zero.
However, as the sensor would be operated at elevated temperatures one must consider the thermal expansion due to the change in temperature:

$$\Delta L = \alpha L_o \Delta T$$ [30] [12]

Where α is the thermal expansion coefficient. Aluminum has the highest expansion coefficient of all materials used in this design. As it does not cover the whole plate it would not generate a dominant effect. If one assumes an effective expansion coefficient of 3 times the largest and thickest layer, the n-epilayer, then the change in length would be 0.36 um in the y direction and 0.15 um in the x direction. Such sub percentage changes are negligible. If one wished to take all possible precautions to prevent buckling one could corrugate the silicon nitride guard ring. There is literature on this topic[31]. One possible corrugation scheme would create ten 4 um deep and 25 um long “grooves” in the silicon nitride ring. The only changes in the process flow occur after the deposition of the n-epi, when the “grooves” are etched. This can be done either isotropically or anisotropically[32].

[1] *Reduction of Malodor by Oral Cleansing Procedures*, J. Tonezetich, S.K. Ng, Oral Surg., August 1976, pp. 172-181.

[2] *Cadaverine as a Putative Component of Oral Malodor*, S. Goldberg, A. Kozlovsky, D. Gordon, M. Rosenberg, J. Dent. Res. 73(6): June 1994, pp. 1168-1171.
[3] *Halitosis Measurement by an Industrial Sulphide Monitor*, M. Rosenberg et alt., J. Peridontol., August 1991, pp. 487-489.
[4] *Ultrasonic Plate Waves for Biochemical Measurements*, B.J. Costello, B.A. Martin, R.M. White, 1989 Ultrasonics Symposium, pp. 977-981.
[5] *A Multisensor Employing an Ultrasonic Lamb-Wave Oscillator*, S.W. Wenzel, R.M. White, IEEE Trans. Electron Devices, vol. 35, no. 6, June, 1988, pp. 735-743
[6] *Flexural Plate-Wave Gravimetric Chemical Sensor*, S.W. Wenzel, R.M. White, Sensors and Actuators, A21-23 (1990), pp. 700-703.
[7] *Surface Acoustic Wave Gas Sensor Based on Film Conductivity Changes*, A.J. Ricco, S.J. Martin, T.E. Zipperian, Sensors and Actuators, 8 (1985), pp. 319-333.
[8] *Stability, Sensitivity and Selectivity of Tungsten Trioxide Films for Sensing Applications*, D.J. Smith, J.F. Vetelino, R.S. Falconer, E.L. Wittman, Sensors and Actuators B, 13-14 (1993), pp. 264-268.
[9] *Surface Chemistry of H2S sensitive Tungsten oxide Films*, B. Fruehberger, M. Grunze, D.J. Dwyer, Sensors and Actuators B31 (1996) pp. 167-174.
[10] *Piezoelectric Properties of Zinc Oxide Films on Glass Substrates Deposited by RF-Magnetron-Mode Electron Cyclotron Resonance Sputtering System*, M. Kadota, M. Minakata, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol 42, No. 3, May 1995, pp.345-349
[11] *Surface Acoustic Wave Devices and Their Signal Processing Applications*, Colin Campbell, Academic Press, ISBN 0-12-157345-1, 1989, p. 17.
[12] *Piezoelectric Properties of ZnO Films Deposited by an ECR Sputtering System*, M. Kadota, T. Kasanami, Electronics and Communications in Japan, Part 3, Vol. 76, No. 11, 1993, pp. 1-9.
[13] *Integrated Circuit Compatible Design and Technology of Acoustic Wave Based Microsensors*, M.J. Vellekoop, G.W. Lubking, P.M. Sarro, A. Venema, Sensors and Actuators A44 (1994) pp. 249-263.
[14] *No2 Gas-Concentraaion Measurement with a SAW Chemosensor*, A. Venema, M. Vellekoop, M. Nieuwenhuizen et. alt, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol UFFC-34, No. 2, March 1987, pp.148-154.
[15] *Surface Wave Devices for Signal Processing*, D.P. Morgan, Elsevier Publishinc Company, ISBN 0-444-42511-X, 1985, chapter 1.
[16] See reference #6.
[17] *Cmos Compatible Silicon Devices on Thin SiO2 membranes*. M.A. Gajda, H. Ahmed, J. Dodson, Electronic Letters, 6th. January 1994, Vol. 30, No. 1, pp. 28-29. The doping concentration should be at least 5 X 10^15 atoms/cm^3.
[18] *A Silicon-Silicon Nitride Membrane Fabrication Process for Smart Thermal Sensors*, P.M. Sarro, A.W. Herwaarden, W. von der Vlist, Sensors and Actuators A, 41-42 (1994) pp. 666-673.
[19] *A High Sensitivity CMOS Gas Flow Sensor on a Thin Dielectric Membrane*, D. Moser, H. Baltes, Sensors and Actuators A, 37-38 (1993), pp. 33-37.
[20] All material properties were found in the following reference books:

CRC Handbook of Chemistry and Physics, 1996.

Comprehensive Inorganic Chemistry, ed. J.C. Bailor, Pergamon Press, 1973.

Hints on how to calculate effective thermal resistive values can be found in:

Electronic Packaging and Interconnection Handbook, C.A. Harper, ed. McGraw Hill Publishing, 1991, chapter 2.

[21] See reference 13# .
[22] *Design Aspects of SAW Gas Sensors*, A. Vennema, E. Nieuwkoop, M.J. Vellekoop, Sensors and Actuators, 10 (1986) pp. 47-64.
[23] See reference #13.
[24] See reference #13.
[25] *Study of Electrochemical Etch-Stop for High Precision Thickness Control of Silicon Membranes*, B. Kloeck, S.D. Collins, N.F. DeRooiji, R.L. Smith, IEEE Transactions on Electron Devices, Vol. 36, No.4, April 1989, pp. 663-669.

[26] *Hydrogen Sulfide Surface Acoustic Wave Gas Detector*, J.F. Vetelino, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol. UFFc-34, No. 2, March 1987, pp. 156-161.
[27] *Stability, sensitivity and selectivity of tungsten trioxide films for sensing applications*, D.J. Smith, J.F. Vetelino, R.S Falconer, E.L. Wittman, Sensors and Actuators B, 13-14 (1993), pp. 264-268
[28] See reference #8.
[29] *Buckling Behavior of Boron doped P+ Silicon Diaphragms*, X. Den, W.H. Ko, Case Western Reserve University, Cleveland Ohio, IEEE 1991, pp. 201-204.
[30] *Surface Micromachined Linear Thermal Microactuator*, J.W. Judy, T. Tamagawa, D.L. Polla, Department of Electrical Engineering, Institute of Technology, University of Minesota, IEEE 1990 pp. 629-632.
[31] *The Fabrication and Use of Micromachined Corrugated Silicon Diaphragms*, J. H. Jerman, Sensors and Actuators, A21-A23 (1990) pp. 988-992.
[32] *Fabrication and Dynamic Testing of Electrostatic Actuators with p+ Silicon Diaphragms*, E.H. Yang, S.S. Yang, S.W. Han, S.Y. Kim, Sensors and Actuators A50 (1995) pp. 151-156.