

Factors affecting the performance of micromachined sensors based on Fabry–Perot interferometry

E Jesper Eklund and Andrei M Shkel

Microsystems Laboratory, University of California, Irvine, Engineering Gateway 2110, Irvine, CA 92697, USA

E-mail: eeklund@uci.edu and ashkel@uci.edu

Received 16 March 2005, in final form 15 June 2005

Published 21 July 2005

Online at stacks.iop.org/JMM/15/1770

Abstract

The noise and performance limitations of optical sensors that utilize Fabry–Perot interferometry detection are investigated. A Fabry–Perot interferometer consists of two partially transparent parallel plates with reflective inner surfaces. The plates form a cavity with an optical resonance that depends on the distance between them. At resonant wavelengths, the incident light energy is transmitted through the sensor and intensity peaks occur. The distance between the plates can be obtained by detecting the wavelength of the transmitted light. Various sensors can be based on Fabry–Perot interferometry, e.g. accelerometers, pressure sensors, and microphones. This paper considers factors affecting the performance of this type of sensor, including mechanical–thermal noise, contribution of noise in the detection system, and effects of reflectivity, surface roughness and parallelism of the mirrors. The presented experimental data support the results of the analysis.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

A Fabry–Perot interferometer (FPI) consists of two parallel mirrors, defining a cavity with an optical resonance that depends on the distance between the mirrors. At resonance, the incident light energy is transmitted through the FPI, forming transmission intensity peaks at the resonant wavelengths. The distance between the plates can be obtained by detecting the wavelength of the transmitted light. In addition to many applications in optical communications, a wide range of sensors can be based on FPI principles. An FPI with a suspended mirror can be used as the sensing element of accelerometers, microphones and pressure sensors.

Detection systems that utilize optics allow for highly sensitive sensors, potentially detecting displacements on a sub-nanometer scale. Furthermore, a passive optical sensor does not require an electrical signal and is therefore highly resistant to electromagnetic interference (EMI) and radio frequency interference (RFI), making these types of sensors well suited for usage in harsh environments. In addition, FPI-based

sensors can be serialized in a linear network by connecting multiple passive sensors between a central light source and a single detection system [1]. These devices are also promising candidates for biomedical applications when direct contact with living tissue is necessary, since no electrical signal is required in the sensor node.

Along with the advantages offered by FPI-based optical detection schemes, there are many design tradeoffs. Some of the issues that need to be considered are mechanical–thermal noise, electrical noise in the detection system, effects of reflectivity, surface roughness and parallelism of the mirrors, assembly, and packaging. The quality of the antireflective coating and the quality of the suspension system are other factors that can potentially degrade the performance of an FPI sensor. This work explores the design considerations and tradeoffs of sensors that are based on Fabry–Perot interferometry. The theoretical performance limitations are compared to experimental results obtained from a micromachined FPI prototype, which is shown conceptually in figure 1.

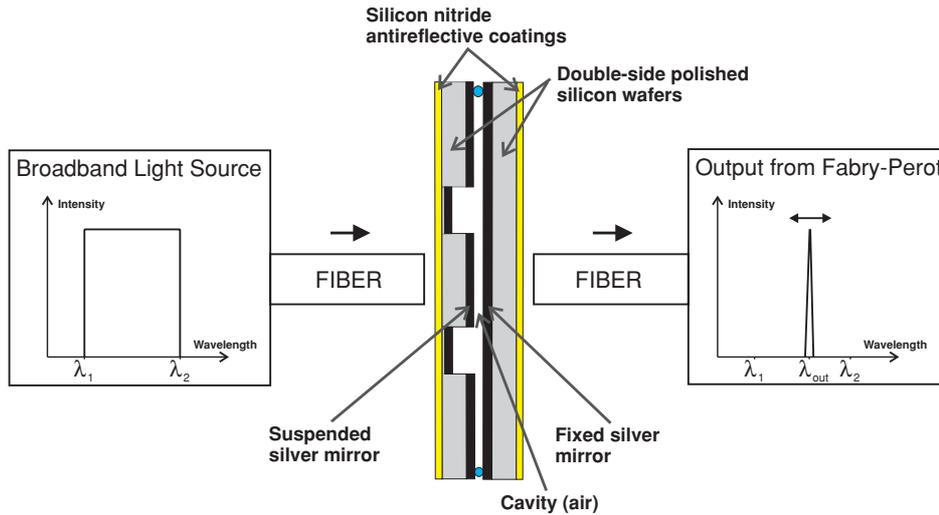


Figure 1. Fabry–Perot interferometer sensor concept.

The principle of operation of sensors based on Fabry–Perot interferometry is described in section 2, followed by noise modeling and other design considerations in section 3. Section 4 presents the fabrication and characterization of a prototype, including measurements of the mirror quality, antireflective coating properties and transmission curves.

2. FPI principle of operation

Two parallel plates with reflective inner surfaces form a Fabry–Perot interferometer. If light with a wide wavelength spectrum is transmitted through the FPI, the intensity of the transmitted light at a particular wavelength depends on the distance between the plates, d , according to the Airy function,

$$\tau = \frac{T^2}{(1-R)^2} \left[1 + \frac{4R}{(1-R)^2} \sin^2 \left(\frac{2\pi nd \cos \theta}{\lambda} \right) \right]^{-1}, \quad (1)$$

where R is the reflectivity of the mirrors, T is the transmission, n is the refractive index, θ is the angle of incidence, and λ is the wavelength [2]. As can be noted from equation (1), the maximum intensity will occur when $\sin^2(2\pi nd \cos \theta / \lambda) = 0$. The plate distance that will generate an intensity peak can therefore be obtained from

$$d = \frac{m\lambda}{2n \cos \theta}, \quad (2)$$

where m specifies the order of interference. By considering consecutive orders of interference in equation (2), the distance between interference peaks, or the free spectral range (FSR), can be calculated as

$$\text{FSR} = \frac{\lambda}{m}. \quad (3)$$

In order to measure the deflection of a sensor, e.g. due to acceleration or pressure, a relationship between deflection and wavelength is derived. By differentiating equation (2) and assuming that the refractive index of the medium remains constant, the deflection can be written as $\Delta d = \frac{d_0}{\lambda_0} \Delta \lambda$, where d_0 is the initial cavity gap, λ_0 is the wavelength that generates a peak at d_0 , Δd is the change in plate spacing due to acceleration or pressure, and $\Delta \lambda$ is the shift in wavelength of the intensity

peak due to Δd . For example, the response of a quasi-static accelerometer is $a = \omega_n^2 \Delta d$, where ω_n is the natural frequency. A direct relationship between the measured acceleration and the shift in wavelength of the intensity peak, $\Delta \lambda$, can now be obtained as

$$a = \frac{m\omega_n^2}{2n \cos \theta} \Delta \lambda. \quad (4)$$

Several different detection schemes can be considered in order to detect the wavelengths transmitted through the Fabry–Perot interferometer. In order to achieve an approximately linear output with a system consisting of a laser diode and a single photodiode only a small part of the transmission curve can be used. This limits the possible deflection of the proof mass and requires the initial plate spacing to be located where the slope of the transmission curve is the steepest, which in turn reduces the range of the sensor. An alternative way of detecting the intensity peak is to sweep the wavelengths of a tunable laser. After the signal is transmitted through the FPI, a photodiode is used to detect the peak. However, due to the limited sweeping rate of tunable lasers, this system can only operate at low frequencies. A third alternative is to use a broadband light source with a wide wavelength spectrum. Once the signal is transmitted through the FPI, the light is optically demodulated with a diffraction grating and directed onto a linear photodiode array. The position where the light hits the array can then be related to the wavelength and in turn acceleration or pressure.

3. Noise and design considerations

Factors that affect the sensor performance include finesse, fabrication imperfections, mechanical–thermal noise, photodetector noise and mechanical damping. These issues are considered and modeled in this section.

3.1. Finesse and fabrication imperfections

The quality of a Fabry–Perot interferometer can be determined by its finesse, which is defined as the free spectral range divided

by the full width half maximum (FWHM) of the interference peaks [2]. In other words, the finesse is a measurement of how many peaks at FWHM the FPI can be tuned over without encountering the next-order fringe and consequently cause an ambiguous output. The finesse depends solely on the reflectivity of the mirrors in an ideal FPI:

$$N_R = \frac{\pi\sqrt{R}}{1-R}. \quad (5)$$

However, fabrication imperfections will reduce the finesse of Fabry–Perot interferometers. These effects were described in [2], where spherical bowing δt_S , surface roughness (root-mean-square deviation) δt_G , and departure from parallelism δt_P were considered. If all three defects are included, the overall defect finesse N_D , at a particular wavelength λ , is determined by

$$\frac{1}{N_D^2} = \left(\frac{2\delta t_S}{\lambda}\right)^2 + \left(\frac{4.7\delta t_G}{\lambda}\right)^2 + \left(\frac{\sqrt{3}\delta t_P}{\lambda}\right)^2. \quad (6)$$

If the defects are considered together with the reflection finesse, an effective finesse, N_E , can be defined by

$$\frac{1}{N_E^2} = \frac{1}{N_D^2} + \frac{1}{N_R^2}. \quad (7)$$

When highly reflective mirrors are used in a Fabry–Perot interferometer, $N_R \gg N_D$, and the effective finesse is mainly limited by the defects in the fabricated device.

3.2. Damping and frequency response

The squeeze-film damping between an FPI-based sensor's fixed and suspended plates can be estimated from

$$b = 0.42 \frac{\mu A^2}{d^3}, \quad (8)$$

where A is the area of the proof mass, μ is the viscosity, and d is the distance between the mirrors [3].

A more intuitive form of the damping is the damping ratio $\xi = b/(2M\omega_n)$, where M is the mass. By utilizing this and assuming that the sensor is modeled as a second-order system, the dimensionless frequency response is

$$|H(\omega)| = \frac{1}{\sqrt{\left[1 - \left(\frac{\omega}{\omega_n}\right)^2\right]^2 + (2\xi \frac{\omega}{\omega_n})^2}}. \quad (9)$$

If the system is undamped ($\xi \approx 0$), the frequency response can be simplified as $|H(\omega)| \approx 1/|1 - (\omega/\omega_n)^2|$. This puts a limit on the usable frequencies of the sensor. For example, if a 5% deviation is acceptable, the maximum operational frequency of an undamped device is $\omega_{\max} \approx 0.22\omega_n$. However, if the system is critically damped ($\xi \approx 0.7$) equation (9) yields an operational frequency as high as $\omega_{\max} \approx 0.59\omega_n$ for the same 5% deviation. Thus, it is desirable to choose the dimensions of the sensor such that critical damping can be obtained. For example, if a cylindrical proof mass with radius 200 μm is defined in a 500 μm thick silicon wafer ($\rho \approx 2330 \text{ kg m}^{-3}$), the gap between the FPI plates required to achieve a critically damped system is $d = [0.42\mu A^2/(2M\omega_n\xi)]^{1/3} \approx 2.4 \mu\text{m}$, assuming operation in air ($\mu \approx 1.8 \times 10^{-5} \text{ N s m}^{-2}$) and a resonant frequency of 7 kHz. From equation (2), the order of interference that will achieve this plate spacing in air at a wavelength of 1550 nm is $m \approx 3$, which in turn yields an FSR of 520 nm from equation (3).

3.3. Mechanical–thermal noise

Mechanical–thermal noise can be calculated using either the equipartition theorem or the Nyquist relation [4], which is a physical equivalent of the thermal noise in electrical resistance. For example, consider the response of an accelerometer $|H(f)|$, which leads to a Nyquist noise displacement of

$$\frac{|X_n(f)|}{\sqrt{\Delta f}} = \sqrt{4k_B T b} \frac{|H(f)|}{k}, \quad (10)$$

where k_B is Boltzmann's constant, T is the temperature, b is the damping, and k is the equivalent spring constant. Following [4], the signal response of an accelerometer can be written as $|X_s(f)| = (f/f_n)^2 |H(f)| |Y_s|$, where $\omega^2 |Y_s|$ is the magnitude of the input acceleration in meters per square second per root hertz. The signal response is therefore

$$\frac{|X_s(f)|}{\sqrt{\Delta f}} = \frac{a |H(f)|}{\omega_n^2}. \quad (11)$$

The signal-to-noise ratio (SNR) of the displacement is now calculated,

$$\left|\frac{X_s}{X_n}\right|^2 = \frac{a^2 M}{8k_B T \xi \omega_n}, \quad (12)$$

where b has been replaced by the dimensionless damping ratio ξ .

By setting $|X_s/X_n| = 1$, the smallest detectable acceleration can be obtained. If the dimensions of the critically damped system in the previous section are used, the noise equivalent resolution is calculated to be 8 μG . Better resolution can be obtained if an underdamped accelerometer is designed by vacuum packaging the device or increasing the plate spacing. However, increased spacing yields a higher order of interference. This reduces the shift in wavelength caused by acceleration and requires a more accurate detection system. Furthermore, as determined in the previous section, the frequency needs to be below $0.22\omega_n$ in an undamped system, which can be compared to the maximum frequency of $0.59\omega_n$ in critically damped accelerometers.

3.4. Photodetector noise

Assuming that a PIN photodiode is utilized, the signal-to-noise ratio (SNR) of the detection system is obtained from

$$\text{SNR} = \frac{(\rho\tau P_{\text{in}})^2}{2q(\rho\tau P_{\text{in}} + I_D) + (4k_B T/R_L)F_n + (\rho\tau P_{\text{in}}r_1^2)\Delta f} \frac{1}{\Delta f}, \quad (13)$$

where q is the electron charge, I_D is the dark current, k_B is Boltzmann's constant, T is the temperature, R_L is the load resistance, ρ is the responsivity of the photodiode, and r_1^2 is a measurement of the noise level in the incident optical signal. The power that is transmitted to the photodiode, P_{pd} , is the product of the power of the light source, P_{in} , and the transmission coefficient, τ , from equation (1). Equation (13) was obtained by considering the photodiode current $I_P = \rho P_{\text{pd}}$, shot noise $i_S^2 = 2q(I_P + I_D)\Delta f$, thermal noise $i_T^2 = (4k_B T/R_L)\Delta f$, the amplifier noise factor F_n , and the relative intensity noise (RIN) of a laser source $i_{\text{RIN}}^2 = I_P^2 r_1^2$ [5].

For example, assume that the responsivity of a photodiode is 1 A W^{-1} , the dark current is $1 \mu\text{A}$, the load resistance is 50Ω ,

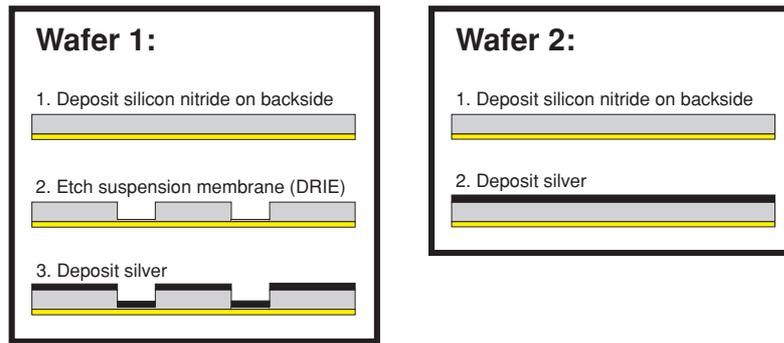


Figure 2. Fabry-Perot interferometer process flow.

the RIN is $r_1^2 = -120$ dB, and the absorption in the FPI and the thermal noise in the pre-amplifier are ignored. For these values, the shot noise and RIN are much smaller than the thermal noise. The signal-to-noise ratio will therefore be limited mainly by the thermal noise. By setting $\text{SNR} = 1$, the noise equivalent power of the photodetector is calculated to be $18 \text{ pW Hz}^{-1/2}$. If the light source is an ideal tunable laser and the wavelength is placed at the maximum slope of the transmission curve of the FPI, the minimum noise displacement can be calculated from

$$\frac{\Delta P_{\text{pd}}}{\Delta d_{\text{min}}} \approx \left[\frac{\partial P_{\text{pd}}}{\partial d} \right]_{\text{max}} = P_{\text{in}} \left[\frac{\partial \tau}{\partial d} \right]_{\text{max}}. \quad (14)$$

The noise displacement is calculated to be $0.6 \text{ fm Hz}^{-1/2}$ for a reflectivity of 0.994, input power 0.1 mW, negligible absorption, zero angle of incidence, operation in air, and the third order of interference. For a natural frequency of 7 kHz this translates to an accelerometer resolution of $0.1 \text{ } \mu\text{G Hz}^{-1/2}$. This is the best possible theoretical resolution, which is only obtainable for a very limited range. Note that these calculations only apply if an ideal laser and a photodiode are used. If a broadband light source and a photodiode array are utilized instead, the resolution is limited by the finesse, diffraction grating, and the number of elements in the photodiode array.

3.5. Antireflective coating

An antireflective (AR) coating is required to reduce the interferometric behavior of the wafer itself. Without the antireflective coating, the sensor will be highly reflective at the transition between air and silicon, which in effect will create three Fabry-Perot interferometers in series. The material used for the coating can be determined from the refractive indices of the transmitting media. Ideally, for a single-layer AR coating, the refractive index of the antireflective material should be the square root of the product of the refractive indices of two materials that it separates [6]. A layer of silicon nitride is utilized in the FPI prototype due to its index of refraction ($n_{\text{N}} \approx 2.0$), which works well as an AR coating between air and single-crystal silicon ($n_{\text{Si}} \approx 3.5$) [7]. To achieve minimum reflection, the optical thickness (thickness of the medium multiplied by its refractive index) of a single-layer AR coating should be a quarter of the wavelength of the light. This yields a thickness of

approximately $2000 \text{ } \text{\AA}$ for silicon nitride and wavelengths around 1550 nm . Experimental results evaluating the single-layer antireflective coating are presented in section 4.4.

4. Experimental results

An FPI prototype, which can potentially be used as an accelerometer, is fabricated from $450 \text{ } \mu\text{m}$ thick double-side polished silicon wafers by defining a membrane suspension with deep reactive ion etching (DRIE). Silver mirrors are deposited using e-beam evaporation.

4.1. Process flow and fabricated parts

The fabrication process is summarized in figure 2. First, a $2000 \text{ } \text{\AA}$ thick PECVD silicon nitride antireflective coating is deposited on the reverse side of the polished wafer at a rate of $144 \text{ } \text{\AA s}^{-1}$ at $250 \text{ } ^\circ\text{C}$ and 17 mTorr pressure. One of the wafers is patterned with AZ4620 photoresist and dry-etched by DRIE until a thin membrane is obtained around the cylindrical proof mass. The continuous DRIE process consists of 360 etch cycles with 14 s etch steps at 700 W and 7 s passivation steps at 600 W , and is completed in 2 h and 6 min at 0.1 mTorr base pressure. A $400 \text{ } \text{\AA}$ thick silver mirror is then deposited on top of both wafer pieces using e-beam evaporation at a deposition rate of $0.5 \text{ } \text{\AA s}^{-1}$ in $4 \text{ } \mu\text{Torr}$ vacuum.

Figure 3(a) shows the fabricated wafer pieces. In order to simplify the alignment and coupling of light through the FPI, a fairly large device is fabricated. The proof mass, which was defined by DRIE, has a diameter of 3 mm and the trench is 2 mm wide. These dimensions can be reduced by at least an order of magnitude and are mainly limited by the sensitivity and accuracy of the alignment stage. Figure 3(b) shows that the bottom of the trench has a very rough surface. This is due to the aggressive recipe used when etching the trench.

4.2. The mirror surface

The mirror surface roughness is measured with an optical surface profilometer (Wyko NT1100). The average peak-to-valley distance after the silver mirrors are deposited is measured to be $64 \text{ } \text{\AA}$, which is comparable to the initial silicon roughness of $18 \text{ } \text{\AA}$ specified by the wafer manufacturer.

Equation (6) can now be used to calculate the defect finesse. Disregarding deviations from parallelism and

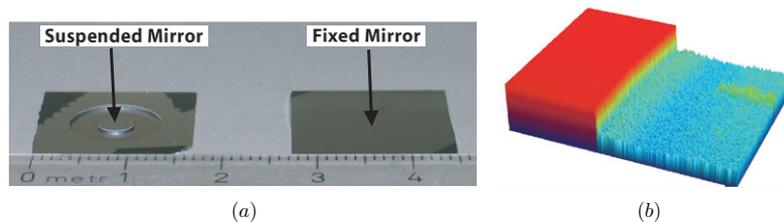


Figure 3. (a) Fabricated wafer pieces, and (b) roughness of the suspension membrane defined by DRIE.

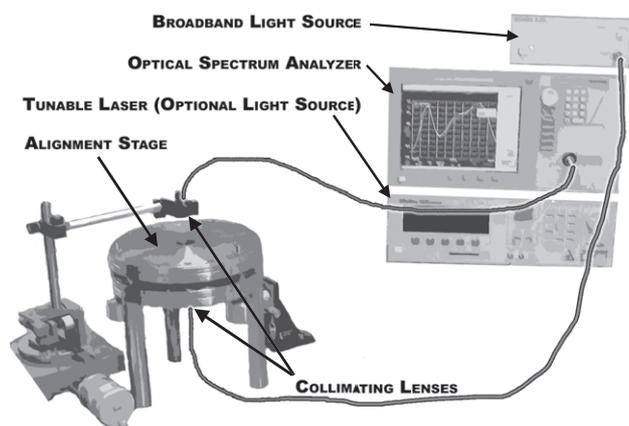


Figure 4. The experimental setup.

spherical bowing, the defect finesse due to surface roughness is $N_D = \lambda / (4.7\delta t_G) = 1550 / (4.7 \times 3.2) \approx 103$. Assuming that the silver mirrors have a reflectivity of 0.994 [8], the reflection finesse is $N_R = \pi\sqrt{R}/(1-R) \approx 522$ according to equation (5). The effective finesse can be obtained from equation (7) and is calculated to be $N_E = 1/\sqrt{1/N_D^2 + 1/N_R^2} \approx 101$.

4.3. The experimental setup

The two wafer pieces are placed in an alignment stage with three degrees of freedom and thin spacers are put between the wafers to reduce the effects of structural vibrations when aligning the wafers. By adjusting the distance between the plates, the FPI can be tuned. A collimating lens is used to focus the light before it is transmitted through the device. At the opposite side of the FPI, another collimating lens is placed to receive the light. A single-mode optical fiber then transmits the signal to an optical spectrum analyzer, which is used to plot the intensity curves. Figure 4 illustrates the experimental setup.

Slow-curing epoxy is applied between the wafers and the alignment stage is adjusted until a maximum value of the transmitted intensity is obtained, indicating that the FPI plates are well aligned. If desired, the spacing between the plates can be decreased to achieve a lower order of interference before the epoxy is cured. To simplify the alignment process, the Fabry–Perot cavity can potentially be defined when fabricating the wafers using either a wet anisotropic etchant or by depositing spacers, e.g. metal, silicon or silicon oxide.

4.4. Undesired wafer fringes

Interference peaks occur as soon as a single silver-coated wafer is placed on the alignment stage, as shown in figure 5(a). These undesired peaks arise from the wafer itself, which functions as a low-finesse Fabry–Perot interferometer. The free spectral range is about 0.7 nm, which yields an order of interference of $m = \lambda/\text{FSR} = 1540/0.7 = 2200$ according to equation (3). The Fabry–Perot cavity length is calculated from equation (2) as $d = m\lambda/2n$ (assuming a light path perpendicular to the surface). Since the cavity now consists of a silicon wafer, the index of refraction is $n = 3.5$ [7]. The wavelength $\lambda = 1540$ nm is obtained from the transmission curve in figure 5(a), which gives $d = 2200 \times 1540 \text{ nm} / (2 \times 3.5) \approx 480 \mu\text{m}$. This can be compared to the specified wafer thickness between 448 μm and 472 μm , thus supporting the notion that the peaks are caused by the wafer itself.

The effects of the undesired interference peaks, caused by the wafer, can be reduced with a layer of silicon nitride antireflective coating (section 3.5). Figure 5(b) demonstrates two different silicon nitride thicknesses. The bottom curve was obtained with a 3750 Å thick silicon nitride layer and the top curve with a 2000 Å layer. This plot confirms that a thickness of 2000 Å gives a low reflectivity, as was calculated in section 3.5. These peaks can potentially be further reduced if a multi-layer antireflective coating is used instead of the single layer of silicon nitride.

4.5. The assembled FPI

Figure 6(a) shows the assembled FPI. A transmission spectrum with a free spectral range of approximately 29 nm is obtained with a tunable laser and displayed in figure 6(b). Due to the low resolution and the oversweeping of the spectrum analyzer, these peaks have been averaged and therefore look smoother than expected. However, this plot is readily used to obtain the FSR of the system.

In order to see the true shape of the peaks, a broadband light source is used and the spectrum analyzer is set to display a narrow wavelength spectrum around the second peak. The FWHM of the peaks is determined to be 0.41 nm from figure 6(c). Note that imperfections in the antireflective coating (figure 5(b)) lead to two small sidebands in addition to the transmitted peak. A multi-layer coating can potentially be used to eliminate these sidebands. From equation (3), an order of interference of $m = \lambda/\text{FSR} = 1540/29 \approx 53$ is calculated, which means that the distance between the plates is approximately $d = m\lambda/2n = 53 \times 1540 \text{ nm} / (2 \times 1) \approx 41 \mu\text{m}$ according to equation (2), assuming that the light path is perpendicular to the surface. The natural frequency

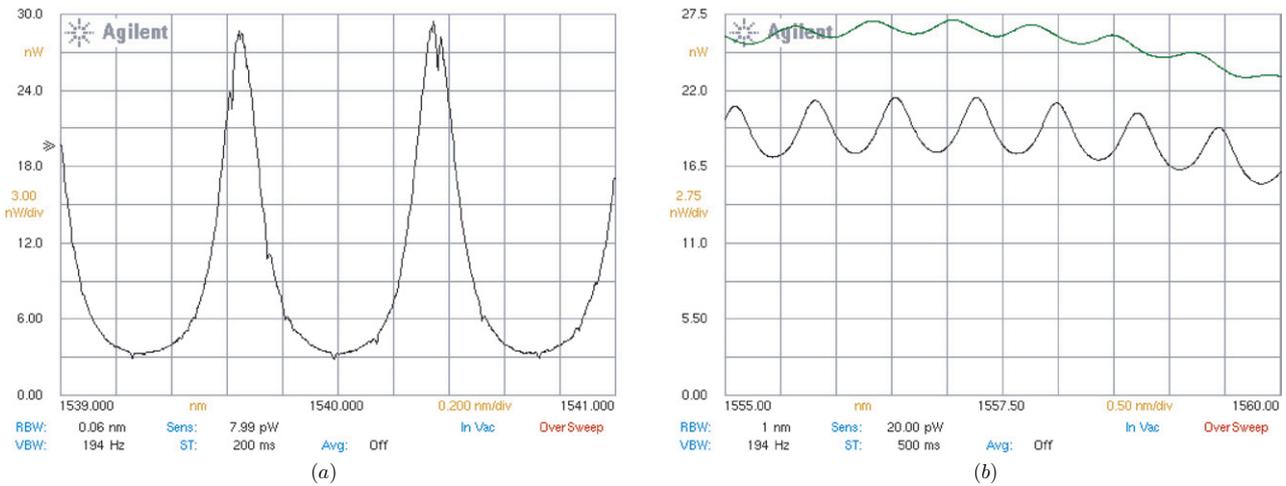


Figure 5. (a) A single silver-coated wafer placed in the path of light behaves like a low-finesse FPI, and (b) antireflective coatings of 2000 Å and 3750 Å silicon nitride are used to reduce this effect.

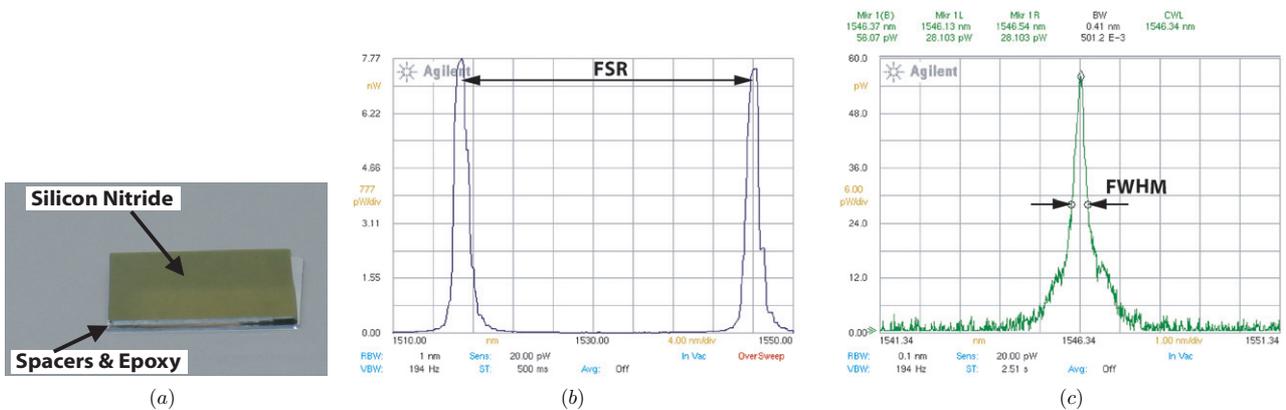


Figure 6. (a) The assembled FPI prototype and (b) its transmission curve. (c) FWHM of the transmission curve in (b).

of the device was estimated to be 1350 Hz, yielding an accelerometer resolution of 80 mG from equation (4). Better resolution can be achieved if the natural frequency is lowered, very flat high-reflectivity dielectric mirrors are used, and the order of interference is reduced by decreasing the cavity gap.

Defined as the FSR divided by the FWHM of the interference peaks (section 3.1), a finesse of $N = 29/0.41 \approx 71$ is obtained. This can be compared to the estimated finesse of 101, calculated in section 4.2. A deviation from parallelism likely accounts for the difference between the measured and estimated values. Assuming this is the only cause, the departure from parallelism is calculated to be 8.9 nm according to equation (6).

5. Conclusion

The performance of optical sensors based on Fabry–Perot interferometry was discussed. These types of sensors are small and robust, highly resistant to EMI and RFI, and can potentially detect displacements on a sub-nanometer scale. It was demonstrated how the performance of FPI-based sensors depends on mechanical–thermal noise, photodetector noise,

fabrication imperfections and assembly. The main limitations are the surface quality and alignment of the mirrors. Surface roughness, curvature or a slight deviation from parallelism will greatly reduce the finesse and in turn the resolution of the sensor. Experimental results from a membrane-suspended FPI with silver mirrors showed that the finesse decreased by almost an order of magnitude due to surface roughness and deviation from parallelism.

Acknowledgments

This research was partially supported by UC Discovery/VIP Sensors grant no 9-442531-19919-8 and NSF grant no CMS-0330470.

References

[1] Perez M and Shkel A 2005 Conceptual design and preliminary characterization of serial array system of high-resolution MEMS accelerometers with embedded optical detection *Smart Struct. Syst.* **1** 63–85

- [2] Atherton P, Reay N, Ring J and Hicks T 1981 Tunable Fabry–Perot filters *Opt. Eng.* **20** 806–14
- [3] Andrews M, Harris I and Turner G 1993 A comparison of squeeze-film theory with measurements *Sensors Actuators A* **36** 79–87
- [4] Gabrielson T 1993 Mechanical–thermal noise in micromachined acoustic and vibration sensors *IEEE Trans. Electron Devices* **40** 903–9
- [5] Agrawal G 2002 *Fiber-Optic Communication Systems* (New York: Wiley) pp 156–70
- [6] Hass G and Thun R 1964 *Physics of Thin Films* vol 2 (New York: Academic) pp 243–4
- [7] Weber M 2002 *Handbook of Optical Materials* (Boca Raton, FL: CRC Press)
- [8] Bennett J and Ashley E 1965 Infrared reflectance and emittance of silver and gold evaporated in ultrahigh vacuum *Appl. Opt.* **4** 221–4