A Monolithic Approach to Fabricating Low-Cost, Millimeter-Scale Multi-Axis Force Sensors for Minimally-Invasive Surgery

Joshua B. Gafford¹, Student Member, IEEE, Samuel B. Kesner², Member, IEEE, Alperen Degirmenci¹, Student Member, IEEE, Robert J. Wood³, Member, IEEE, Robert D. Howe³, Fellow, IEEE, and Conor J. Walsh³, Member, IEEE

Abstract—In this paper we have rapidly prototyped customized, highly-sensitive, mm-scale multi-axis force sensors for medical applications. Using a composite laminate batch fabrication process with biocompatible constituent materials, we have fabricated a fully-integrated, 10x10 mm three-axis force sensor with up to 5 V/N sensitivity and RMS noise on the order of ~1.6 mN, operational over a range of -500 to 500 mN in the x- and y-axes, and -2.5 to 2.5 N in the z-axis. Custom foil-based strain sensors were fabricated in parallel with the mechanical structure, obviating the need for post-manufacturing alignment and assembly. The sensor and its custom-fabricated signal conditioning circuitry fit within a 1x1x2 cm volume to realize a fully-integrated force transduction platform with potential haptics and control applications in minimally-invasive surgical tools. The form factor, biocompatibility, and cost of the sensor and signal conditioning makes this method ideal for rapid-prototyping low-cost, mm-scale distal force sensors. Sensor performance is validated in a simulated tissue palpation task using a robotic master-slave platform.

I. INTRODUCTION

The advent of minimally-invasive surgical (MIS) techniques has initiated a paradigm shift in surgery. Procedures once requiring a large incision, resulting in significant morbidity and recovery times, can now be performed through a discrete number of millimeter-sized ports which can be quickly stitched or patched. Robotic MIS surgery in particular is experiencing increasing commercialization and clinical adaptation for a limited subset of procedures. However, in light of substantial advancements in robotic control and dexterity, a haptic chasm still separates the surgeon from his or her anatomical workspace. When interacting with delicate anatomy, the lack of haptic feedback can lead to numerous complications including intraoperative hemorrhage, tissue damage, and suture breaking [1].

Haptic feedback has been explored in many areas of surgery, including laparoscopic surgery, microsurgery, and vitreoretinal surgery [2]. Haptic feedback implementations can be broken into kinesthetic (force sensing) and cutaneous (tactile sensing) modalities [3]. Although cutaneous implementations necessitate distal sensing to accurately reconstruct the tactile profile, such a requirement is relaxed for kinesthetic sensing where only the magnitude of the applied force is desired. Coupling these two modalities can offer a full reconstruction of the applied force over the regime of interest.

Numerous examples of force transduction elements for haptics and MIS exist in literature. These elements range from optical FBG modalities [4]–[7], microfabricated PZT [8], semiconductor-based strain gages [9], and soft microchannels filled with liquid metal [10] to name a few. These approaches often suffer from prohibitively high costs of manufacturing and assembly, exceedingly complicated or expensive signal conditioning infrastructures, susceptibility to thermal drift, limited range of applied forces, and limitations on achievable linearity and resolution. In addition, due to strict size constraints imposed by MIS, it is often the case that the force sensors are placed proximally or remotely with respect to the tool/end-effector. As such, measured forces at the tool tip are contaminated by friction and reaction forces at the point of entry, as well as actuation forces and the mechanics of the tool itself. Thus, there is a desire to develop a new class of cost-effective, mm-scale sensors that can be implemented distally to generate a pure measurement of the applied force at the tissue interface. Herein we present a novel approach to batch fabrication of low-cost, high-resolution three-axis force sensors, as shown in Figure 1. The sensor is constructed out of biocompatible materials and fits within the size constraints imposed by MIS procedures. Manufactured using a process inspired by printed circuit
board manufacturing (printed-circuit MEMS, or PCMEMS) [11], [12], the mechanical structure, metallic thin foil sensing elements, and signal conditioning PCB are all fabricated in one integrated, monolithic manufacturing process. This approach facilitates high-throughput batch fabrication of fully-integrated sensors in parallel, significantly driving down costs and lead times associated with manufacturing and assembly. Proper arrangement of passive and active hinges allows for out-of-plane features to stiffen the overall structure and localize deformation to the sensitive elements. Such a sensor is intended for applications in which the applied force is relatively low (over a range of 1-2 N), including tissue palpation and haptic feedback in microsurgery to name a few.

II. ANALYSIS AND OPTIMIZATION

To leverage the capabilities of the PCMEMS process, we have adopted a thin-beam ‘Maltese Cross’ topology consisting of a suspended platform (‘probe’) attached to structural ground via axisymmetric flexures that transmit motion in the desired sensing degrees of freedom. This morphology is shown in Figure 2 (left). Signals generated by strain gages bonded to these flexures can be combined and algorithmically decoupled via a calibration matrix to determine the magnitude and direction of the force applied at the probe in N axes so long as $n_{gages} \geq N$. Herein we present a 3-axis sensor, where each beam has two gages (one in tension and one in compression). The combination of signals generated by each beam under load can be reconstructed into a 3-dimensional applied force vector.

A. Strain Theory

Loading conditions experienced by the sensor are shown in Figure 2 (right). When a load in the $z$-direction is applied axially to the probe tip, each flexure is in a state of pure bending with a vertical displacement $u(x_b)$ (where $x_b = 0$ at the attachment to the structural ground). The resulting surface strain $\varepsilon(z,l)$, given as a function of position $x_b$ along each beam, is described by Equation 1:

$$\varepsilon(z,l)(x_b) = -\frac{F_x h}{8E I} \left( x_b - \frac{l_b}{2} \right)$$

where $F_x$ is the applied load, $h$ is the distance between strain and the beam’s neutral axis, $E$ is the Young’s Modulus, $I$ is the second moment of area, and $l_b$ is the length of the beam.

When a load is applied in the $x$- or $y$-direction, a moment about the central platform produces an angular displacement $\theta$ at the platform as given by Equation 2 [13]. The resulting surface strain $\varepsilon(x_l)$ is given by Equation 3.

$$\theta = \frac{F_x L}{2} \left( \frac{E I r}{I_0} \left( r + \frac{l_b}{2} \right) + \frac{E I r}{I_0} \left( r + \frac{l_b}{3} \right) + k_t \right)$$

$$\varepsilon(x_l)(x_b) = \frac{6h \theta}{l_b^2} \left[ \left( r \left( \frac{x_b}{l_b} - \frac{1}{2} \right) + l_b \left( \frac{x_b}{2l_b} - \frac{1}{3} \right) \right) \right]$$

where $F_x$ is the applied load, $r$ is the distance defined in Figure 2 (right), and $L$ is the probe length.

These strain profiles, normalized and plotted as a function of distance along the beam (also normalized to beam length $l_b$), are shown in Figure 2 (right) for each loading condition. In the case of $z$-loading, the strain profile is antisymmetric about the midpoint of the beam, whereas for $x$- or $y$-loading, the zero strain crossing is slightly offset from the midpoint.

At millimeter scales, gage coverage area comprises a significant portion of the beam surface area, so it is inaccurate to consider only the maximum strain in gage analysis. Since the strain is obviously linear over the length of each beam, it is adequate to consider the average value of strain integrated over gage length. Thus, we define the average strain $\overline{\varepsilon}$ as follows:

$$\overline{\varepsilon} = \frac{1}{l_{g,f} - l_{g,i}} \int_{l_{g,i}}^{l_{g,f}} \varepsilon(x_b) dx_b$$

where $l_{g,i}$ and $l_{g,f}$ are the gage’s start and end position along length $x_b$ of the beam.

As we are primarily interested in average strain, we limit the gage length to half of the beam length (noting that if the gage length exceeds half of the beam length, desensitization occurs due to a reversal in the direction of strain as shown in the normalized strain plots in Figure 2 (right)).

B. Methodology for Designing Custom Gages

Using the PCMEMS fabrication process, we have the ability to custom-design metal foil strain gages on an application-
From a sensitivity standpoint, making the gage as short as possible is ideal. However as we diminish the characteristic gage length, the effects of Joule heating become substantial \( Q_{\text{dis}} \propto R_s^{-1} \), resulting in thermal expansion. Although the geometric axisymmetry of the mechanical structure helps to cancel out some thermal expansion, significant gradients could still compromise sensor stability in terms of thermal drifting.

C. Flexure Beam Geometry Optimization

The previously-derived analysis was used to guide a brute-force optimization study to determine beam geometry where the objective function is to maximize sensitivity \( S \) (defined here as \( V_{\text{out}}/F_{\text{app}} \)) subject to the mechanical constraint that the sensor does not yield for an applied force of 1N. As the length of the beams are roughly constrained by the outer dimensions of the sensor (10mm x 10mm), beam width \( w \) and beam thickness \( t \) were chosen as free parameters in the optimization. The results of such an analysis for \( x \) and \( y \) loading is shown in Figure 4, where we have chosen a design sensitivity of approximately 5 V/N (satisfied by \( w = 1.25 \text{mm} \) and \( t = 0.20 \text{mm} \)) given a 15mm long probe. Note that this corresponds to a mechanical factor of safety (defined as \( \sigma_y/\sigma_{\text{max}} \)) of around 2.5 as shown by the contour lines. These dimensions result in a \( z \)-sensitivity (analysis not shown) of roughly 0.5 V/N.

III. SENSOR DESIGN AND FABRICATION

The mechanical structure of the sensor is fabricated in parallel with the sensing elements to realize a fully-integrated multi-axis sensor with no need for post-manufacturing alignment, bonding or assembly. Precision laser machining and alignment during lamination ensure dimensional tolerances on the order of 10 \( \mu \text{m} \) for manufacturing repeatability.

A. Manufacturing

The sensor is comprised of a multi-material laminate consisting of several functional sub-laminates, as shown in Figure 5 (top). 304 Stainless Steel shim stock (4 layers,
each layer is 50 µm thick) forms the structural substrate which will dominate the mechanical behavior of the sensor. Kapton polyimide (25 µm thick) is used as (1) a flexible layer allowing active hinges to transmit motion, and (2) an insulating and encapsulating layer to isolate the gage. Constantan (55% Cu / 45% Ni, 5 µm thick) is used as the gage material due to its versatility and similarity to 304SS in terms of thermal expansion. DuPont Pyralux FR1500 sheet adhesive (12.5 µm thick) is used to bond subsequent layers. All materials with the exception of the adhesive are biocompatible; a thin Parylene coating can be deposited onto the sensor structure for complete biocompatibility.

The manufacturing process of the sensor is similar to that outlined in [14]. Layers are individually machined using a diode-pumped solid-state (DPSS) laser, plasma etched with Argon gas to promote surface adhesion, and laminated in a heat press at 50 psi and 200°C for 2 hours. After lamination, the sensor gage pattern is cut using the laser to realize a nominal resistance of 100 Ω, and the sensor is encapsulated in an additional Kapton layer (with breakout contacts) prior to the final release cuts. After release cuts are made, the stiffening struts are folded, locked, and soldered, and the sensor is wired to the signal conditioning circuitry.

Various assembly, wiring, and folding features are machined into the laminate, as shown in Figure 5 (bottom). A combination of active hinges (castellated hinges approximating pin-joint motion), plastic hinges (serrated material that plastically deforms when folded), and snap-fit features allow for precision folding into a robust, 3-dimensional assembly which stiffens the structure and localizes deformation to the flexural elements. Microscope images of the sensor prototype, pre-encapsulation, are shown in Figure 6.

B. Finite Element Analysis

Although we can use the previously-derived analytical model to approximate the strain behavior of the sensor, the underlying assumption is that the flexure beam attachment points are themselves perfectly rigid. Commercial off-the-shelf (COTS) sensors typically satisfy this assumption by substantially increasing material thickness at the attachment. Using PCMEMS fabrication, where we approximate structural rigidity using plastic hinges and features that fold out-of-plane and lock into place, the elastic behavior becomes more complicated at these interfaces. As such, finite element analyses (Solidworks Simulation, Dassault Systems), shown in Figure 7, were performed to (1) qualitatively determine the location of maximum (average) strain across each beam (to determine the optimum gage placement), and (2) quantitatively validate the previously-derived analytical model used in optimization.

Note that the beams are attached at corners for increased...
stiffness. As the FEA results elucidate in Figure 7, this attachment point is not perfectly rigid and a small amount of local deformation occurs at the corners. Thus it becomes beneficial to place the gages on the probe-half of the beam for optimum sensitivity, where the attachment of the probe more closely approximates a rigid connection point.

C. Signal Conditioning

An on-board signal conditioning circuit was designed to amplify the gage output such that reasonable TTL voltage levels can be processed by a DAQ. On-board sensing ensures that resistive/inductive contributions from leadwires are minimized pre-amplification, so the design challenge is to minimize the footprint of the signal conditioning PCB to similar dimensions as the sensor itself.

As shown in Figure 8, a half bridge with precision (0.05% tolerance) reference resistors feeds an AD8221 instrument amplifier with tunable gain. The amplifier gain was set to 1000 such that a 500 mN load results in ∼2.5 V voltage swing. The half-bridge configuration ensures adequate temperature compensation and long-term stability. A nodal analysis was performed to select a high-impedance tuning potentiometer which allows for manual zero-offset calibration coverage for up to 5% manufacturing mismatch in gage resistance with negligible effect on linearity and sensitivity. The sensor is designed for a single supply, so the input-output relationship is highly linear. To obtain the calibration matrix to convert from signal to source, we can compute the Moore-Penrose pseudoinverse to our overconstrained system of equations where \( \text{dim}(s) \succ \text{dim}(f) \) and \( \text{rank}(C^T) = \text{dim}(f) \) which can be shown physically [15]–[17]. The Moore-Penrose approach presents a common static decoupling technique that entails the use of a least-squares algorithm to numerically compute the pseudoinverse of the calibration matrix \( C \), which is given in Equation 12.

![Signal conditioning circuit schematic](image)

\[ f = [C]_{13 \times 4} s \]  
\[ s = [-5.33 -0.65 0.59] \]  
\[ 5.11 0.03 0.57 \]  
\[ 0.00 -4.53 0.50 \]  
\[ 0.25 5.64 0.59 \] (11)

Note that the sensor is significantly more sensitive in the \( x \)- and \( y \)-axes than the \( z \)-axis due to the mechanical amplification of the probe. Sensitivity matching can be achieved simply by shortening the probe length which results in less mechanical amplification of \( x \)- and \( y \)-forces and allows the sensor to be placed more distally. As is evident from the calibration data given in Figure 9, the input-output relationship is highly linear. To obtain the calibration matrix to convert from signal to source, we can compute the Moore-Penrose pseudoinverse to our overconstrained system of equations where \( \text{dim}(s) > \text{dim}(f) \) and \( \text{rank}(C^T) = \text{dim}(f) \) which can be shown physically [15]–[17]. The Moore-Penrose approach presents a common static decoupling technique that entails the use of a least-squares algorithm to numerically compute the pseudoinverse of the calibration matrix \( C \), which is given in Equation 12.

\[ f = \begin{bmatrix} -0.096 & -0.096 & 0.009 & -0.004 \ 0.010 & -0.015 & -0.093 & 0.102 \ 0.463 & -0.463 & 0.447 & 0.408 \end{bmatrix} s \] (12)

Sample performance data for the \( x \)-axis is given in Table I. The sensor is designed to withstand mechanical loads of up to 2 N so the gain can simply be adjusted to accommodate a higher operational range given the 5V supply. The RMS noise of the sensor, measured by integrating the power spectral density of a null signal, is roughly 8mV (corresponding to 1.6mN in the \( x \)- and \( y \)-axis). Note that, due to increased stiffness, the RMS noise is a factor of 10 higher in the \( z \)-direction (roughly 16 mN).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
<th>% Full-Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>5.1</td>
<td>V/N</td>
<td>N/A</td>
</tr>
<tr>
<td>Range</td>
<td>±500</td>
<td>mN</td>
<td>N/A</td>
</tr>
<tr>
<td>Linearity</td>
<td>2.0</td>
<td>mN</td>
<td>0.20</td>
</tr>
<tr>
<td>Noise</td>
<td>1.6</td>
<td>mN RMS</td>
<td>0.16</td>
</tr>
<tr>
<td>Hysteresis</td>
<td>10</td>
<td>mN</td>
<td>1.00</td>
</tr>
</tbody>
</table>
behavior is accurately mapped by a second-order model of the form $\ddot{x} + 2\zeta\omega_n^2 x + \omega_n^2 x = 0$, where $\zeta = \beta/2\sqrt{mk}$ is the damping coefficient, $\omega_n = \sqrt{k/m}$ is the natural frequency, $k$ is the analytical stiffness of the sensor, $m$ is the combined mass of the sensor and probe, and $\beta$ is damping added by the adhesive and polyimide layers (empirically determined). An interesting property is that the sensor is almost critically damped ($\zeta = 0.88$). The measured 10%–90% rise time $t_r$ is roughly 18 ms, limiting the bandwidth $f_{-3dB}$ to about 20 Hz (using a first-order approximation for bandwidth given by $f_{-3dB} = 0.35/t_r$). Since there is a tradeoff between bandwidth ($f_{-3dB} \propto t^2$) and sensitivity ($S \propto t^{-2}$), increasing the thickness of the steel substrate could greatly improve the dynamic range at the cost of sensitivity.

C. Tissue Palpation Simulation

A tissue palpation experiment was simulated wherein the sensor was attached to the end-effector of a 3-DoF linear stage equipped with a 3-DoF wrist to realize a 6-DoF micromanipulation platform. An image of the test setup is shown in Figure 10 (left). Three ball-screw linear stages (ATS100-100, Aerotech, Pittsburgh, PA) are mounted orthogonally to create the 3-DoF linear stage. The linear stages have a 100 mm travel with a 0.5 μm resolution and are connected to a control box (A3200 Npaq Drive Rack, Aerotech, Pittsburgh, PA) that runs an internal servo loop on the stages at 8 kHz. The 3-DoF wrist has a gimbal design with three direct drive rotary joints. Three 12 Watt 22-mm brushless DC motors (EC-max 283840, Maxon, Switzerland) are used to actuate the wrist. The motors are controlled using three digital positioning controllers (EPOS2 24/2, Maxon, Switzerland) that run an internal servo loop at 1 kHz. The micromanipulation stage is controlled through a GUI (Graphical User Interface) that handles communications with the Aerotech and Maxon motors through provided API (Application Programming Interface) library calls. The GUI was programmed in C++ using the Qt application framework (Digia Plc, Helsinki, Finland).

A tissue analog was molded out of Ecoflex-0010 (Smooth-On), with high-stiffness intrusions (steel balls) embedded at different depths to simulate metastatic or cancerous tissue regimes. As a preliminary quantitative evaluation, a solid block of Ecoflex-0010 was palpated with a spherical sensor probe. The force profile generated by the sensor was compared to an analytical force profile based on (1) a simple model assuming linear-elastic Hertzian contact mechanics [18], and (2) a hyperelastic, neo-Hookean Mooney-Rivlin model [19] given an empirical Young’s Modulus of $E = 30$ kPa for the Ecoflex-0010. The individual force components, as well as the stiffness curve, are shown in Figure 10 (middle) and (right). The sensor accurately captures the nonlinear hyperelastic behavior of the elastomer, as well as the hysteretic behavior that is not captured in the model.

For the palpation task, the tissue analog was probed in an 8x8 grid, as shown in Figure 11 (left), with each subsequent probe separated from the previous probe by 9.2 mm. Steel balls were buried at depths of 1 mm (ball 1), 3 mm (ball 2), and 5 mm (ball 3). The indentation speed was 1 mm/s, with a nominal indentation depth of 2mm.

A normalized stiffness contour plot of the tissue as measured by the sensor is shown in Figure 11 (middle) and (right). As can be seen, the sensor was able to accurately
reconstruct the stiffness map of the tissue and localize high-stiffness intrusions with the exception of the deepest intrusion which was 5 mm deep (most likely due to the shallow indentation depth).

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have demonstrated novel monolithic fabrication of a fully-integrated, highly-sensitive triaxial sensor using printed-circuit MEMS techniques. Both analytical and numerical tools were employed in optimization analyses used to determine the sensor’s geometric parameters. Design and fabrication rules were generated for manufacturing a compact and temperature-stable signal conditioning circuit. The performance of the sensor was tested both in benchtop calibration experiments as well as a simulated tissue palpation task to demonstrate the sensor’s efficacy at evaluating unknown forces with high precision. Preliminary characterization methods have shown that the sensor can resolve forces with ≃ mN resolution at a frequency of 20 Hz. Additionally, the sensor was able to accurately detect subtle stiffness changes in a simulated tissue.

Future work will focus on refining the manufacturing process and downsizing the size of the sensor without significantly compromising performance. Alternative gage manufacturing methods, such as sputtering, will also be investigated to increase nominal resistances. Structural enhancements will be implemented to improve the bandwidth over 100 Hz.

ACKNOWLEDGEMENT

The authors would like to acknowledge the Wyss Institute for Biologically-Inspired Engineering and the Harvard School of Engineering and Applied Sciences for their support of this work, and Dr. Frank L. Hammond III for his help with the micromanipulation stage.

REFERENCES


