A High-Force, High-Stroke Distal Robotic Add-On for Endoscopy*

Joshua B. Gafford, Student Member, IEEE, Robert J. Wood, Senior Member, IEEE, and Conor J. Walsh, Member, IEEE

Abstract—"Snap-On" robotic modules that can integrate distally with existing commercially-available endoscopic equipment have the potential to provide new capabilities such as enhanced dexterity, bilateral manipulation and feedback sensing with minimal disruption of the current clinical workflow. However, the desire for fully-distal integration of sensors and actuators and the resulting form factor requirements preclude the use of many off-the-shelf actuators capable of generating the relevant strokes and forces required to interact with tools and tissue. In this work, we investigate the use of millimeter-scale, optimally-packed helical shape memory alloy (SMA) actuators in an antagonistic configuration to provide distal actuation without the need for a continuous mechanical coupling to proximal, off-board actuation packages to realize a truly plug-and-play solution. Using phenomenological modeling, we design and fabricate antagonistic helical SMA pairs and implement them in an at-scale roboendoscopic module to generate strokes and forces necessary for deflecting tools passed through the endoscope working port, thereby providing a controllable robotic 'wrist' inside the body to otherwise passive flexible tools. Bandwidth is drastically improved through the integration of targeted fluid cooling. The integrated system can generate maximum lateral forces of 10N and demonstrates an additional 96 degrees of distal angulation, expanding the reachable workspace of tools passed through a standard endoscope.

I. INTRODUCTION

Endoscope-based systems enable intraluminal access to confined anatomy through natural orifices, resulting in both cosmetic and economic benefits over more invasive procedures [1]. While there is growing interest in performing complex therapeutic procedures endoscopically, such as endoscopic submucosal dissection (ESD), endoscope technology itself has remained relatively unchanged for several decades, and as a result, widespread applications thereof are primarily diagnostic in nature. There is a real incentive to innovate in the field and close the gap between the state of the technology and the demand for more complex minimally-invasive procedures.

A. Prior Work

Robotic endoscopy is a rich academic field encompassing a number of areas, including but not limited to teleoperation using novel actuation systems, haptic sensing, novel actuation technologies, rapid manufacturing of highly-articulate continuum structures, capsular approaches, and the design and control of dexterous robotic end-effectors [2]–[8]. Commercially, companies like Auris Robotics, Medrobotics and Invendo Medical are beginning to spin out these technologies into commercial products that promise greater access to and visualization of anatomy through natural orifices.

A characteristic common among numerous endoscopic robots presented in literature is the intimate mechanical coupling between distal end-effectors and proximal actuator systems. Typically this coupling is achieved via tendon-sheath systems (TSS). This approach has resulted in significant innovation in the field of flexible surgical robotics, granting endoscopists with unprecedented dexterity in vivo. The requirement for flexibility can result in hysteretic and nonlinear behaviors arising from tendon compliance, friction, and capstan effects resulting from sheath curvature. These
effects make such systems difficult to model and control. In addition, proximal actuator packages that drive these systems can be large, cumbersome and potentially occlusive to the surgical arena, thereby requiring alterations to the clinical workflow to allow for docking/registration and undocking of these systems.

Due to fundamental limitations in manufacturing of mm-scale electromechanical systems and the dearth of commercially-available sensing and actuation options that conform to the strict form factor and robustness requirements, very few groups have focused on developing affordable, distal robotic modules with fully-integrated sensors and actuators [9]. We are working to develop robotic systems with sensors and actuators located entirely at the distal end to remove the mechanical coupling and realize a truly modular and deployable system. This work leverages prior work in design and monolithic fabrication of mechanisms, sensors and actuators at the mm-scale [10]–[12].

B. Contribution

Deployable endoscopic robotic systems, as shown in Fig. 1, preclude the use of conventional actuators (DC motors, voice coils, etc.) due to size limitations and the desire to remove the mechanical coupling between distal and proximal ends. As such, with the goal of ‘moving everything to the tip,’ we consider alternative actuation strategies that are acquiescent to the rigorous form factor requirements of fully-distal implementations. One such strategy employs the use of energy-dense phase changing materials (shape memory alloys, or SMAs) to provide low-to-moderate actuation speed and high force for actuating distal structures. In our previous work [10] we have shown that wire-based SMA actuators can be used for distal actuation of an articulating structure. However, the low strain capabilities (4-5%) ultimately limited the dexterity we were able to demonstrate.

To enable greater dexterity while matching (or exceeding) force output capabilities of previous designs, we require higher-stroke actuation systems capable of generating a high net lateral force to deflect common flexible endoscopic tools. To this end, we explore the use of helical SMA actuators. With the objective functions of (a) minimizing the overall form-factor (spring diameter and length) while (b) maximizing the force and stroke, the problem can be naturally posed as an optimization of work per unit volume. However, given the difficulties associated with accurately modeling helical SMA actuators, it can be deceitful to rely on model-based techniques alone to find a local optimum. As such, we implement an approach that combines approximate phenomenological modeling and experimental validation to converge on and identify a configuration that meets functional requirements. After discovering an optimized configuration, we implement the actuators in an at-scale robotic module and demonstrate greatly improved stroke characteristics.

In Section II we present a fixed-strain constitutive model used to approximate the blocking force evolution, as well as a quasi-steady version of the full phenomenological antagonism model for predicting net force vs. strain. In Section III we discuss the fabrication and control of helical SMA antagonistic pairs per model requirements. In Section IV we present experimental results of the optimized actuator design and validate both models. Section V discusses the integration of the actuators into an at-scale roboendoscopic module analog and preliminary results thereof.

II. SMA Modeling and Fabrication

Several SMA-based flexible robotic systems with medical applications have been demonstrated in literature [13], [14]. However, these systems typically suffer from limited force generation [15], limited stroke [16], or prohibitively slow actuation speeds [17]. In this work, we aim to generate high-force, high-stroke actuation at practical speeds through heuristic design of optimum actuator configurations coupled with the integration of forced fluidic cooling.

A. Fixed-Strain Model for Blocking Force

Modeling the blocking force of helical SMA actuators involves solving the quasi-steady, fixed strain form of the shear-based constitutive equation [18], [19]:

$$
\tau - \tau_0 = G_s(\gamma - \gamma_0) + \Omega_S(\xi_S - \xi_S^0) + \Theta(T - T_0)
$$

(1)

where $\tau$, $\tau_0$ is instantaneous/initial shear stress, $\gamma$, $\gamma_0$ is the instantaneous/initial shear strain, $\xi_S$, $\xi_S^0$ is the instantaneous/initial martensitic fraction, $T$, $T_0$ is the instantaneous/initial temperature, $G_s$ is the shear modulus, $\Theta$ is the thermal expansion factor, and $\Omega_S$ is the stress-induced phase transformation contribution factor (which can be expressed as $\Omega_S = -\gamma_L G_T$ where $\gamma_L$ is the maximum residual strain) [20]. Note that, for blocking force (fixed strain), the first term cancels out of Equation (1). Using the dual-phase model ($\xi = \xi_S + \xi_T$ is the sum of the shear- and temperature-dependent martensitic fractions), the shear modulus is phase-dependent:

$$
G_s = G_a + \xi(m - G_a)
$$

(2)

where $m$ and $G_a$ are the shear moduli in the martensitic and austenitic phases, respectively. The instantaneous martensitic phase components are dependent on initial conditions and the overall phase fraction as follows:

$$
\xi_S = \frac{\xi_S^0}{\xi_0} (\xi_0 - \xi)
$$

(3)

$$
\xi_T = \frac{\xi_T^0}{\xi_0} (\xi_0 - \xi)
$$

(4)

The phase kinetics can be described by a cosine model proposed by [18], where the total martensitic fraction for the reverse transformation (martensite to austenite) is given by:

$$
\xi(T, \tau) = \frac{\xi_0}{2} \left[ \cos \left( a_A \left( T - A_s - \frac{\sqrt{3} T |\tau|}{C_A} \right) \right) + 1 \right]
$$

(5)
where $\xi_0$ is the initial martensite fraction ($\xi_0 = 1$), and $\alpha_A = \frac{\pi}{A_s - A_f}$ where $A_s$ and $A_f$ are the austenite start and finish temperatures respectively, and $C_A$ is a curve-fit parameter.

In anticipation of current-based control, we describe the temperature dynamics using a simple convective heat transfer model modified to include the effects of Joule heating due to an applied current (note: this assumes that there are no thermal gradients, radial or axial, within the wire):

$$mC_p \dot{T} = I^2 R - h c A_c (T - T_\infty)$$

where $m$ is the mass of the actuator, $C_p$ is the specific heat, $I$ is the applied current, $R$ is the actuator’s electrical resistance, $h_c$ is the convective heat transfer coefficient, $A_c$ is the actuator’s surface area, and $T_\infty$ is the ambient temperature. Finally, from a geometric perspective, we can transform initial pre-stretch $\delta_0$ into shear strain $\gamma_0 = \gamma(\delta_0)$ via the following [21]:

$$\gamma(\delta_0) = \frac{3d_w}{4D_s} \cos(\alpha_i) (\sin(\alpha_f(\delta_0)) - \sin(\alpha_i)$$

where relevant geometric parameters are given in Fig. 2, and initial and final pitch angles $\alpha_i$ and $\alpha_f$ are expressed as:

$$\alpha_i = \tan^{-1} \left( \frac{d_w}{2D_s} \right)$$

$$\alpha_f(\delta_0) = \sin^{-1} \left( \frac{\delta_0 \cos(\alpha_i)}{\pi N D_s} + \sin(\alpha_i) \right)$$

A blocking force analysis requires the following set of initial conditions: $\{T_\infty, \tau_0, \xi_{S0}, \gamma_0\}$. A pre-stretch $\delta_0 \rightarrow \gamma_0$ while in the fully martensitic state will result in induced shear stress $\tau_0$ and, depending on the amount of stress, some initial martensitic detwinning $\xi_{S0}$. We note that, although temperature $T$ can be solved for explicitly, equations relating shear stress $\tau$, stress-induced phase volume $\xi_S$, and temperature $T$ are transcendentally related, requiring the use of gradient-based techniques to solve for the stress and phase kinetics for a given temperature. An alternative, more computationally tractable approach is to exploit the slow phase dynamics and solve for the equations in discrete time by ‘updating’ the phase and stress at each point along the temperature profile, and initializing the process by solving for some initial configuration $\tau_0 = G_m \gamma_0 - G_m \xi_{S0} + \Theta T_\infty$ and $\xi_{S0} = \xi(T_\infty, \tau_0)$ simultaneously using gradient descent or a nonlinear solver. As a heat profile is applied, we can evolve

$$\xi_S^{(n+1)} = \begin{cases} \xi(T^{(n)}, \tau^{(n)}) \xi_S^{(n)} & \text{if } A_s < T^{(n)} < A_f \\ \xi_S^{(n)} & \text{otherwise} \end{cases}$$

where $A_s' = (A_s + \sqrt{\xi^{(n)}_S G_{\tau}})$ is the stress-modified austenitic start temperature and $A_f' = (A_f + \sqrt{\xi^{(n)}_S G_{\tau}})$ is the stress-modified austenitic final temperature. The shear modulus and stress are subsequently updated:

$$G_{\tau}^{(n+1)} = G_{\tau} + \xi(T^{(n)}, \tau^{(n)})(G_m - G_{\tau})$$

$$\tau^{(n+1)} = -\gamma_L G_{\tau}^{(n+1)} (\xi_S^{(n+1)} - \xi_{S0}) + \Theta(T^{(n)} - T_\infty) + \tau_0$$

where $\nu$ is the Poisson ratio of the material.

**B. Modeling Antagonism**

If we wish to model antagonistic behavior with strain dynamics, as shown in Fig. 3, the model becomes more complicated as we can no longer ignore the first term on the right side in Equation (1). In this case we assume that the active actuator is biased by a (geometrically identical) passive actuator in martensitic state where the martensitic de-twinning phenomena are stress-dominated and can be described by the cosine model:
where $\tau_{s}^{cr}$ and $\tau_{f}^{cr}$ are the critical shear stresses at the start and finish of the conversion, respectively, and $\tau_{b}$ is the shear stress in the martensitic state. Note that we introduce subscript $b$ to differentiate the bias spring dynamics from those of the active spring.

As before, we have the following discrete-time update law for the phase kinetics describing the detwinned martensite volume fraction:

$$\xi_{Sb}^{(n+1)} = \begin{cases} \xi_{Sb}^{(n)} - \tau_{f}^{cr} (\tau_{b} - \tau_{s}^{cr}) \cos \left( \frac{\pi}{\tau_{s}^{cr} - \tau_{f}^{cr}} (\tau_{b} - \tau_{s}^{cr}) \right) + \frac{1 + \xi_{Sb}}{2} & \text{if } \tau_{s}^{cr} < \tau_{b}^{(n)} < \tau_{f}^{cr} \\ \xi_{Sb}^{(n+1)} & \text{otherwise} \end{cases}$$ (15)

In contrast to the blocking force model, the strain rate is no longer negligible, and is a function of the system kinematics. Here we assume that the strain rate in the active actuator is dictated by the elastic compliance of the bias element and some net blocking force (which we assume is constant), and we also assume steady-state thermodynamics within the bias element. Casting into discrete time and using a simple force balance shown in Fig. 3, given some instantaneous deflection $\Delta^{(n+1)}$, we can write $F_{b}^{(n+1)}$ and $F_{a}^{(n+1)}$ as:

$$F_{b}^{(n+1)} = K \left( G_{m}, \alpha \delta_{f} (\delta_{0} + \Delta^{(n+1)}) \right) \left[ \delta_{0} + \Delta^{(n+1)} \right]$$

$$- \frac{\pi \delta_{m}^{3}}{8 \kappa_{b}^{2}} G_{m} \gamma_{L}(\xi_{Sb}^{(n+1)} - \xi_{Sb0})$$

$$= F_{a}^{(n+1)} - F_{net}$$ (17)

$$F_{a}^{(n+1)} = K \left( G_{r}^{(n+1)}, \alpha \delta_{f} (\delta_{0} - \Delta^{(n+1)}) \right) \left[ \delta_{0} - \Delta^{(n+1)} \right]$$

$$- \frac{\pi \delta_{m}^{3}}{8 \kappa_{a}^{2}} G_{r} \gamma_{L}(\xi_{Sa}^{(n+1)} - \xi_{Sa0})$$ (18)

We re-iterate that the phase characteristics in the active actuator have contributions from both detwinned martensite and austenite during the heating cycle, whereas the bias actuator phase is characterized by the former only.

We can solve Equations (17) and (18) for $\Delta^{(n+1)}$ using gradient-based techniques to solve for deflection as a function of time given a known input current profile. Alternatively, a much simpler analysis is to prescribe a known set of $\Delta$ and set $G_{r} \rightarrow G_{a}$ and $\xi_{Sa} \rightarrow 0$ in Equation (18) (i.e. putting the active actuator in a fully austenitic state) to observe the force margin (i.e. $F_{a} - F_{b}$) as a function of stroke. This is the approach we implemented later in Section IV when comparing the model with experimental results. Regardless of the approach, the shear strains in the active and bias actuators are updated accordingly: $\gamma_{a}^{(n+1)} = \gamma_{b}^{(n+1)} = \gamma(\Delta^{(n+1)})$. Shear stresses and moduli are incremented as in the previous section for both the active and bias actuator, but with the addition of the strain term.

III. FABRICATION AND CONTROL

A. Fabrication

Custom helical actuators were fabricated in-house using the setup shown in Fig. 4(a). The setup allows for flexible fabrication over a wide parameter space, accommodating a range of different wire diameters and core sizes. A tension following system ensures optimal packing (pitch = $d_{w}$). Fig. 4(b) shows the actuator both in the shape-setting clamps (left), and after annealing at 400C for 1 hour (right).

B. Electronics and Control

1) Power/Sensing Electronics: Custom power/sensing electronics, shown in Fig. 5, were designed and fabricated to enable digital control of the current in two actuators simultaneously via pulse-width modulation (PWM). PWM control was selected due to ease of digital implementation as the inherent low-pass of SMA thermomechanics filters the high-frequency PWM to simulate an analog current control signal. A square wave with varying duty cycle drives the gate of a power NPN MOSFET which sinks current through the SMA, a variable ceramic power resistor that serves as a current limiter, and whatever series resistance arises
to mimic the slow thermomechanical response of the SMA actuator. This filtered signal then drives a PD loop operating at 2 kHz which modulates the duty cycle of the output signal. The controller is implemented in MATLAB xPC and embedded on a PC104 real-time target machine (Diamond Systems Aurora SBC, DMM-32DX-AT Analog/Digital I/O).

IV. VALIDATION

To support actuator characterization efforts, a precision evaluation platform was designed and fabricated to allow for simultaneous SMA current control, execution of position/force-controlled motion profiles, and data logging.

A. Experimental Setup

The experimental setup is shown in Fig. 6. With a design motivated in part by the off-board system developed in [22], the setup features a custom designed validation platform consisting of a moving metrology stage actuated by a preloaded ballscrew drive (Nook Industries) with high-resolution linear encoder (AK Lida 47, Heidenhain, Inc.) feedback. A Maxon EC-Max brushless DC motor with a 14:1 planetary gearhead drives the ballscrew through a custom quasi-back-to-back dual radial bearing system and a timing-belt transmission. A Copley Accelnet ACP-090-09 controller performs the primary-side commutation and closes a low-level current loop. A PC104 xPC target system (consisting of a Diamond Systems Aurora SBC, Diamond Systems Jupiter-MM-LP power supply, Diamond Systems DMM-32DX-AT DAQ, and Sensoray 526 quadrature decoder) implements the outer position/force control loop (PID with gravity compensation and velocity/acceleration feed-forward) and data acquisition/filtering in real time. The target system also provides two independent 50Hz PWM signals to drive the antagonistic SMA power electronics with the current control discussed earlier. Force feedback is provided via an ATI Nano17 titanium 6-axis force sensor. A Dell Precision 5810 Workstation serves as the host computer and communicates with the target via ethernet.

B. Blocking Force Evolution

To validate the thermomechanical blocking force model given in Equation (13), actuators were clamped into the system and pre-stretched by 100% of their unstretched length (i.e. $\delta_0 = L_0$). Varying step currents were applied, and the resulting force was recorded via the ATI Nano17. Example results for the optimized actuator are shown in Fig. 7. The model adequately predicts the blocking force evolution as a function of time and current. We observe that rise time can be arbitrarily controlled via the application of more current, and the dynamics scale as expected.

C. Net Force vs. Stroke (Antagonism)

To determine net force vs. stroke, actuators were tested in tension in both martensitic and austenitic states by applying a triangular displacement profile and measuring the resulting
were stretched to 100% of their unstretched length, and a current of 1.4A was applied to initiate the austenitic phase transition. After steady-state was reached, the current was removed, and the actuators were allowed to cool under three conditions: (a) free convection with air, (b) forced convection with air (via a computer fan), and (c) forced convection with water (by applying droplets of water via a syringe). The results, shown in Fig. 9, demonstrate that the cooling speed is dramatically improved via forced convection techniques. This experiment validates the inclusion of forced fluidic cooling infrastructures to achieve practical actuation speeds.

V. INTEGRATED SYSTEM

To demonstrate practical implementation of the high-force, high-stroke antagonistic actuation scheme, a to-scale robotic module was built to interface with an Olympus CF-100L 13mm endoscope. The module design and preliminary integrated system tests are described below.

A. Module Design

The fabricated module is shown integrated onto the endoscope in Fig. 10(a). An exploded view of the integrated module is shown in Fig. 10(b). The actuator module contains channels for the parallel actuators, as well as in inlet for fluid flow and features for dispersing the flow over the actuators. The actuator module was appropriately sized such that its inclusion does not interfere with the endoscope’s built-in degrees of freedom in the interest of system transparency when not in use. The parallel actuators are coupled to the deflection plate via high temperature Kevlar fiber to fully isolate the SMA from the anatomy with the addition of protective covers which snap into the actuator module. A transparent deflector plate is laser-cut from optically clear acrylic and fastened to a pin-joint-based articulating element. The deflection plate features a through-hole that engages with tools passed through the working channel, thereby transmitting forces to the tool to deflect it laterally.

A torque balance about the pin joint as shown in Fig. 10(c) yields the following expression:

\[ F = \frac{r}{e} \left( r \alpha F - o \right) \]

D. Forced Cooling

A notable drawback of SMA actuation systems is the fact that the speed of actuation is rate-limited by thermodynamic processes. We demonstrated that it is possible to speed the heat-up time by applying more current. To improve cooldown time, we explore the effects of forced convection in different fluid mediums (air and water) to justify the inclusion of a cooling system for performance enhancement. Actuators

\[ F = \frac{r}{e} \left( r \alpha F - o \right) \]
B. Workspace Characterization

To determine the reachable workspace, a 6-axis IMU (MPU6050) was rigidly fixed to the deflector plate, and the system was actuated from negative to positive extreme while angular data was collected along the yaw axis. The results were filtered using both an extended Kalman filter and a complimentary filter (Fig. 11(a)) to show that the system can generate +48/-48 degrees of distal articulation for 96 degrees overall, thereby exceeding our functional requirement of 90 degrees [10]. Fig. 11(b) shows the resulting reachable workspace assuming a rigid transformation between the deflector plate and the tool tip, and the theoretical force profile is computed from Equation (19). As ESD primarily targets lesions with diameters of 2cm and less, and prior studies showed forces on the order of 300-400 mN required to remove tissue during electrosurgery [10], the proposed module is capable of providing the range of articulation and force required to assist the circumferential incision of gastric tumors.

C. Open-Loop Control and Demonstration

An open-loop controller was implemented in MATLAB xPC and embedded on the PC104 system described previously. The controller implements the aforementioned current controller and generates current profiles with high-current transients for fast heat-up which slew to a steady-state low-current to sustain the transition temperature while preventing overheating. Meanwhile, digital logic signals control solenoid valves which provide fluid cooling to the module. This sequence is shown in Fig. 12 along with the resulting yaw angle as measured by the IMU. Observe that fluid cooling allows the system to traverse through its range of motion at...
least 10 times in 55 seconds, in about the same amount of time it would take to switch once without cooling.

The prototype was integrated onto the distal end of an Olympus CF-100L endoscope and into the large intestine of a GI simulator (AK107 Lower GI Endoscopy Simulator, Adam-Rouilly). A flexible GI probe (Conmed) was passed through the endoscope and into the deflector cap, and the module was actuated through its range of motion. Images of this process are shown in Fig. 13. The results demonstrate the clarity of the visual field, as the tip of the tool is always visible. In addition, the system generates enough force to deflect the tool, while still allowing axial translation.

VI. CONCLUSION AND FUTURE WORK

In this paper we present a distal robotic module with integrated actuation capable of generating an additional 96 degrees of deflection and sufficient lateral forces for fine control of endoscopic tools. Using a combination of theory and heuristics, we designed an optimum actuator for generating the required force and stroke, and implemented the actuator in a to-scale modular roboendoscopic system. Future work will focus on design for manufacturing such that the components can be molded, as well as proprioceptive sensor implementation to enable closed-loop control for automated trajectory execution using a similar fabrication approach as described in [10]. In addition, we will collaborate with experimental endoscopists to test the device ex-vivo on appropriate ESD analogs (pig stomachs) as a stepping stone towards in-vivo animal studies on anesthetized pig models.

REFERENCES


