A Si-Micromachined 162-Stage Two-Part Knudsen Pump for On-Chip Vacuum

Seungdo An, Member, IEEE, Naveen K. Gupta, and Yogesh B. Gianchandani, Fellow, IEEE

Abstract—This paper investigates a two-part architecture for a Knudsen vacuum pump with no moving parts. This type of pump exploits the thermal transpiration that results from the free-molecular flow in nonisothermal channels. For a high compression ratio, 162 stages are serially cascaded. The two-part architecture uses 54 stages designed for the pressure range from 760 to $\approx 50$ Torr, and 108 stages designed for lower pressures. This approach provides greater compression ratio and speed than using a uniform design for each stage. Finite element simulations and analytical design analysis are presented. A five-mask single-wafer fabrication process is used for monolithic integration of the Knudsen pump that has a footprint of $12 \times 15$ mm$^2$. The pressure levels of each stage are measured by integrated Pirani gauges. Experimental evaluation shows that, using an input power of $\approx 0.39$ W, the evacuated chamber is reduced from 760 to $\approx 0.9$ Torr, resulting in a compression ratio of $\approx 844$. The vacuum levels are sustained during 37 days of continuous operation. 

Index Terms—Micropump, Knudsen pump, thermal transpiration, two-part, multistage.

I. INTRODUCTION

Knudsen pumps, first proposed and demonstrated in 1909 [1], present an appealing method for obtaining vacuum. These motionless pumps are based on the phenomenon of thermal transpiration [2]. If the hydraulic diameter of a flow channel is no larger than the mean free path of gas molecules (i.e., the flow in the channel is confined to the free-molecular or transitional flow regimes), a thermal gradient induces gas streams from the cold end to the hot end of the channel [2], [3]. Such a pumping method provides advantages over conventional motion-based pumping techniques. First, the absence of moving parts, without frictional loss and mechanical failure, provides significantly higher reliability. Second, this type of pump can potentially have a small form factor [4], [5]. Third, it can provide a high compression ratio using serially-cascaded configuration [1], [6], [7].

Microfabrication technology has been exploited for Knudsen pumps [4], [5], [7]–[14]. Miniaturized pumps are potentially useful for tunable vacuum control system in sealed cavities that contain high-Q resonators, such as micro gyroscopes [15] or timing oscillators [16]. Miniaturized analytical instruments, such as micro mass spectrometers [17], [18] and micro gas chromatographs [19]–[21], also need small vacuum pumps.

A monolithic Si implementation that used one stage was reported in 2005 and demonstrated a compression ratio of $\approx 2$ [4]. The compression ratio can be increased to a limited extent by increasing the operating temperature. However, a more effective and scalable way to increase compression ratio is to cascade stages. Miniaturized multistage Knudsen pumps have been reported in the past [14], [22], but the early efforts were limited in scope because they did not utilize lithographic fabrication. In 2012, a monolithically fabricated 48-stage Knudsen vacuum pump was reported [7], resulting in a high compression ratio of 15 at atmospheric ambient pressure.

Together with the compression ratio, the pumping rate in a multistage Knudsen pump can be increased by adjusting the channel hydraulic diameter for the steady state operating pressure of each stage. An increased hydraulic diameter provides lower hydraulic resistance (i.e., higher pumping rate) as long as gas flow is confined to the free-molecular or transitional flow regime for Knudsen pumping. The Knudsen pump described by Gupta et al. [7] used a uniform design for 48 stages, demonstrating the evacuation of on-chip cavities from 760 Torr to $< 50$ Torr. The hydraulic diameter of the channels in this pump was designed to ensure the free-molecular or transitional flow at atmospheric pressure. For pumping to lower pressures, the hydraulic diameter of the upstream stages must be increased in order to increase the pumping rate. The compression ratio is not sacrificed because in steady state these upstream stages maintain relatively low cavity pressures at which the mean free path is much longer than at atmospheric pressure. However, the larger hydraulic diameter cannot sustain pumping in downstream stages. For this reason, a multistage Knudsen pump intended for a wide pressure range should be partitioned, so that both pumping rate and compression ratio can be enhanced for different parts of the pressure range.

In this context, this paper explores two-part customization of stages for a micromachined Knudsen pump. Section II describes the two-part multistage architecture of a Knudsen pump, the heater design, and the Pirani pressure gauge. Section III details microfabrication processes and results. Test methods and results are presented in Section IV, followed by conclusions in Section V.

1Portions of this work have been published in conference abstract form in [23].
\[ \Delta P_S = \Delta P_f - \Delta P_r \]  

where \( \Delta P_f \) is the unintended, reverse transpiration pressure difference that could be developed in the wide channel at low pressures, at which the flow is not entirely viscous. A serial cascade increases the compression ratio: \( \Delta P_{\text{Total}} = P_{\text{OUT}} - P_{\text{IN}} = \sum \Delta P_{S,i} \), where \( P_{\text{OUT}} \) and \( P_{\text{IN}} \) are the pressures at the OUT and IN ports, respectively, and \( i \) denotes the stage number.

The proposed two-part architecture (Fig. 1) is intended to enhance both pumping rate and compression ratio within an operating pressure range for each part. 1) The Knudsen high-pressure part (KHP), located downstream, is customized for the range from atmospheric pressure to 50 Torr. 2) The Knudsen low-pressure part (KLP), located upstream, is intended for the range from 50 Torr to sub-Torr pressure. The KHP and the KLP are comprised of serially cascaded single-stage Knudsen pumps, using 54 and 108 stages, respectively. The number of stages in the KLP is double that in the KHP, reflecting the larger compression ratio in the KLP than that in the KHP. In principle, the narrow and wide channels for each stage can be designed to provide the highest compression for the intended steady state operating conditions of that stage. However, this amount of dimensional diversity can significantly increase the manufacturing complexity. In this effort, two types of stage designs are pursued, as a compromise.

The following flow equation describes the volume flow rate of gas, \( \dot{V}_X \), from the cold end to the hot end in a long rectangular channel \( [7],[24],[25] \):

\[ \dot{V}_X = \left( \frac{Q_T \Delta T}{T_{\text{Avg}}} - \frac{Q_P \Delta P}{P_{\text{Avg}}} \right) a^2 b^2 \frac{m}{2 k_B T_{\text{Avg}} \rho_{\text{Avg}}} \frac{1}{l} \]  

(2)

where \( a \), \( b \), and \( l \) are, respectively, the height, width, and length of the channel; \( m \) is the mass of a single gas molecule; \( k_B \) is the Boltzmann constant; \( \Delta T \) and \( T_{\text{Avg}} \) are, respectively, the temperature difference and the average temperature of \( T_{\text{Cold}} \) and \( T_{\text{Hot}} \); \( \Delta P \) and \( P_{\text{Avg}} \) are, respectively, the pressure difference and the average pressure of the cold and hot chambers; \( \rho_{\text{Avg}} \) is the mass density at \( P_{\text{Avg}} \) and \( T_{\text{Avg}} \); and \( Q_T \) and \( Q_P \) are, respectively, the thermal creep flow and Poiseuille flow coefficients. Here, \( Q_P \) represents the viscous flow. Note that the terms in \( \Delta T \) and \( \Delta P \) have opposite signs to indicate that the viscous flow balances the thermal creep flow.

The values of \( Q_T \) and \( Q_P \) are obtained by Sharipov’s calculation, which appropriately represents direct simulation Monte Carlo (DSMC) [26]. Sharipov [24] numerically solved the linearized Boltzmann transport equation (BTE), for extracting \( Q_T \) and \( Q_P \) from rarefied gas flow. Each flow coefficient was tabulated for the variations in ratio of \( a \) to \( b \) \((= a/b)\), and the rarefaction parameter. The rarefaction parameter is defined as the inverse of Knudsen number, which is pressure-dependent. For parameters that are unlisted in [24], the values of \( Q_T \) and \( Q_P \) can be determined from interpolations of the tabulated values.

Instead of the volume flow rate, the standard flow rate, \( \dot{V}_{\text{Std}} \), can be used to comprehend molecular flow through a multistage Knudsen pump where each stage is at a different pressure. To obtain \( \dot{V}_{\text{Std}} \), the mass density, \( \rho_{\text{Std}} \), on the right side of Eq. (2), is substituted with the mass density at standard conditions, \( \rho_{\text{Std}} \), provided by:

\[ \rho_{\text{Std}} = \frac{m P_{\text{Std}}}{k_B T_{\text{Std}}} \]  

(3)

where \( P_{\text{Std}} \) is the standard pressure of 760 Torr and \( T_{\text{Std}} \) is the standard temperature of 273.15 K. The standard flow rate allows easy comparison of molecular counts, whereas the volume flow rate is indicative of the swept volume. In Eq. (2),
the subscript \(X\) in \(\dot{V}_X\) denotes non-standard conditions defined by \(P_{\text{Avg}}\) and \(T_{\text{Avg}}\). In a given narrow channel, as pressure decreases, the standard flow rate naturally decreases as well. However, the volume flow rate increases because \(Q_T\) increases relative to \(Q_P\). This is reflected in Eq. (2) and (3).

The pressure difference across a channel, \(\Delta P\), depends upon the flow rate, as indicated in Eq. (2). When the pressure difference is zero (i.e., there is no pressure head), the flow rate takes the largest value. Over time, when evacuating a blind cavity, gas flow approaches equilibrium (i.e., no net flow), where the thermal creep flow due to \(\Delta T\) is completely balanced by the viscous return flow due to \(\Delta P\). The resulting equilibrium pressure difference, \(\Delta P_{\text{Eq}}\), is given by:

\[
\Delta P_{\text{Eq}} = \gamma \frac{\Delta T}{T_{\text{Avg}}} P_{\text{Avg}}
\]

where \(\gamma\) is the flow coefficient ratio, defined as \(Q_T/Q_P\), indicating the relative ratio of the thermal creep flow to Poiseuille flow coefficients in a channel [27]. In the free-molecular regime where the Knudsen number is larger than 10, \(\gamma\) takes its largest value of 0.5 for \(a/b\) ratio of 1 (square) to 0.44 for \(a/b\) of 0 (plate); in the viscous flow regime, where the Knudsen number is smaller than 0.01, \(\gamma\) takes its smallest value of 0 [24]. In the transitional regime, where the Knudsen number is between 10 and 0.1, \(\gamma\) is determined from [24] (Fig. 2a).

Using the determined \(\gamma\) values and Eq. (4), the values of \(\Delta P_f\) for the narrow channel and \(\Delta P_w\) for the wide channel at equilibrium are calculated. Hence, the net pressure difference in a stage, \(\Delta P_{\text{Std}}\), as shown in Eq. (1), is directly proportional to the difference between the \(\gamma\) value for the narrow channel, \(\gamma_n\), and that for the wide channel, \(\gamma_w\). For a large value of \(\gamma_n\), the narrow channel height is preferably no larger than the mean free path. In contrast, to achieve a small value for \(\gamma_w\), the wide channel height is preferably much larger than the mean free path. In narrow channels of a fixed width, the pressure difference and the flow rate show opposite trends to the variation in channel height, \(a\). Typically, a decrease in channel height, \(a\), is accompanied by an increase in forward pressure difference, \(\Delta P_f\), and a decrease in standard flow rate, \(V_{\text{Std}}\), and vice versa.

### B. Selection of Channel Dimensions

Channel dimensions selected for the narrow and wide channels are summarized in Table I. At 760 Torr, a narrow channel height of 0.1 μm in the KHP is similar to the N₂ mean free path of 0.07 μm (Table I); the resulting \(\gamma_n\) value of 0.22 provides a reasonable forward pressure difference in the narrow channel (Fig. 2a). At 50 Torr, the \(\gamma_n\) value in the KHP increases to 0.37. However, at 50 Torr, in the KLP, where the narrow channel height of 1 μm is similar to the mean free path of 1.12 μm, the resulting \(\gamma_n\) value is 0.27, which is smaller than the \(\gamma_n\) value of 0.37 for the KHP at the same pressure. While the loss in the \(\Delta P_f\) due to the diminished \(\gamma_n\) value from 0.37 to 0.27 is small, the standard flow rate of \(6 \times 10^{-6}\) sccm at 50 Torr in the KLP is 30 times larger than \(2 \times 10^{-7}\) sccm at 50 Torr in the KHP and even 3 times larger than \(2 \times 10^{-6}\) sccm at 760 Torr in the KHP (Fig. 2b). The significantly larger value of the standard flow rate in the KLP is due to the quadratic dependence of the flow rate on the channel height, \(a\), as shown in Eq. (2).

In the KHP, the wide channel height is 30 μm, considerably larger than the mean free path of 0.07 μm at 760 Torr; the \(\gamma_w\) value of 0.00 sufficiently suppresses the \(\Delta P_f\) (Fig. 2a). In the KLP, the wide channel height is 100 μm, considerably larger...
than the mean free path of 1.12 μm at 50 Torr. This reduces the magnitude of \( \Delta P_r \), as shown in the plots of \( \gamma_w \) (Fig. 2a). The wide channel height in the KHP is sized smaller than that in the KLP to reduce the dead volume represented by the wide channels, thereby resulting in faster pumping time with a negligible loss in suppressing the \( \Delta P_r \).

In this manner, the two-part design is accomplished by customizing the narrow channels for higher standard flow rate and the wide channels for smaller dead volume, while achieving high compression ratios in both KHP and KLP stages.

C. Calculated Equilibrium Pressures

The equilibrium pressures at each stage for the KHP and the KLP are theoretically calculated using Eq. (4) (Fig. 3). The intended design for the Knudsen pump (Table I) results in the upstream pressures of 65 Torr for the KHP and 0.39 Torr for the KLP, using \( \Delta T \) of 50 K and \( T_{Cold} \) of 300 K (represented by blue solid lines in Fig. 3). All calculations assume that the KHP vents to a downstream pressure of 760 Torr, whereas the KLP has a downstream pressure of 50 Torr. The possible variations in wide channel height for the KLP and \( \Delta T \) for both KHP and KLP are separately calculated as explained below.

First, using the fixed \( \Delta T \) of 50 K, the effects of the increasing wide channel heights from 30 μm to 400 μm for the KLP on pressures are calculated (Fig. 3a). The upstream pressure for the KLP improves from 0.96 Torr for a wide channel height of 30 μm to 0.39 Torr for 100 μm, and further to 0.22 Torr for 400 μm. The hypothetical case in which \( \Delta P_r \) is neglected is also plotted in Fig. 3a. This shows that the \( \Delta P_r \) is increasingly noticeable at sub-Torr pressure as the wide channel height is decreased.

Second, using the fixed channel dimensions listed in Table I, the effects of the increasing values of \( \Delta T \) from 50 K to 150 K on pressures for both KHP and KLP are calculated (Fig. 3b). For the KHP, by increasing values of \( \Delta T \) from 50 K to 100 K and to 150 K, the upstream pressure decreases from 65 Torr to 6 Torr and to 2 Torr, respectively. For the KLP, the upstream pressure decreases from 0.39 Torr to 0.08 Torr and to 0.04 Torr, respectively, but by smaller ratios than those for the KHP. The response for larger \( \Delta T \), especially at sub-Torr pressure, illustrates that the ability to reduce pressure by increasing \( \Delta T \) is neutralized by \( \Delta P_r \).

D. Heater Design and Pirani Gauge

The layout of a single stage in the Knudsen pump is illustrated in Fig. 4. The narrow channel is formed by dielectric thin films located on the surface of the device, whereas the
TABLE II
DOE METHOD FOR LOW POWER HEATER DESIGN. TWELVE DESIGN FACTORS OF ARBITRARY LOW AND HIGH VALUES ARE USED TO FIND $\Delta T$ FOR AN INPUT POWER OF 2.4 mW. THE FOUR MAIN FACTORS, WHICH ARE BOXED, ARE FINE-TUNED FOR FINAL DIMENSIONS

<table>
<thead>
<tr>
<th>Factors</th>
<th>Input 2-level</th>
<th>Output</th>
<th>Final (KHP/KLP)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low (µm)</td>
<td>High  (µm)</td>
<td>$\Delta T$ (K)</td>
</tr>
<tr>
<td>Wide channel</td>
<td>Width</td>
<td>300</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>250</td>
<td>350</td>
</tr>
<tr>
<td></td>
<td>Height</td>
<td>30</td>
<td>200</td>
</tr>
<tr>
<td>Narrow channel</td>
<td>Width</td>
<td>100</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>30</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Height</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Insulator thickness</td>
<td>ONO-1</td>
<td>1.2</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>ONO-3</td>
<td>1.3</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>Al2O3</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Heater</td>
<td>Length</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Width</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Thickness</td>
<td>0.1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

The input power for the Ti/Pt heater is minimized by increasing thermal isolation, using a design of experiments (DOE) method [28]. Using 12-factor, unpaired, 2-level DOE, finite element analysis (FEA) is performed, allowing each structural factor (parameter) to vary arbitrarily. Four primary dependencies are established by evaluating the impact on $\Delta T$. The dimensions are then fine-tuned: 1) the narrow channel length; 2) the thickness of the oxide-nitride-oxide (ONO) layer (ONO-3); 3) the thickness of the Al2O3 layer; and 4) the heater width. For an imposed input power, a temperature increase for KHP can be regarded as similar to that for KLP because impacts of differences in height and width for the fabricated wide channel (described in Section III) are small, as can be seen from Table II. Using the final values in Table II, an input power of 2.4 mW can achieve 331.4 K if the lower surface of the device is held at 300 K (Fig. 5). For the 162-stage pump that is experimentally evaluated in Section IV, the resulting power is 0.39 W.

Integrated Pirani gauges, located adjacent to pump stages at the perimeter of the array, are used to measure the vacuum levels (Fig. 4c). The Pirani gauge converts pressure within a sense gap into a fractional change in electrical resistance of a localized Joule heater; this is caused by a change in the thermal conductance change of the sense gap with pressure [29]. Four Pirani gauges, connected to stages 1, 54, 99, and 162, are selected for representing the pressure distribution in the overall KHP/KLP layout. These gauges are named P1, P54, P99, and P162, respectively. To enhance the sensitivity for a range from 760 Torr to 1 Torr, the sense gap is designed to be ≈1 µm [30]. This sense gap is identical to the narrow channel height in the KLP, so their fabrication steps are identical.

III. FABRICATION

To implement the Knudsen pump, a single-wafer five-mask fabrication process is used (Fig. 6). The important aspects in the channel fabrication are: 1) the narrow channels for the KHP are defined by a thin (sacrificial) polySi layer (Fig. 6b), whereas the narrow channels for the KLP are defined by both thin and thick (sacrificial) polySi layers (Fig. 6b, c); 2) cavities for wide channels are formed by partially etching the bulk silicon (Fig. 6e); and 3) the walls of the narrow channels are composed of stress-relieved oxide-nitride-
oxide (ONO) layers. The (sacrificial) polySi layers are also used for hydraulic connections between KHP and KLP and between Pirani gauges and nearby stages.

The microfabrication process is initiated by deposition and patterning of the low-pressure chemical vapor deposited (LPCVD) first ONO layer (ONO-1) which is later used as a mask for the cavity etch of the single-crystal silicon substrate (Fig. 6a). The process continues with the deposition and patterning of a thick LPCVD polySi layer of $\approx 0.7 \ \mu m$ (Fig. 6b), followed by the deposition and patterning of a thin LPCVD polySi layer of $\approx 0.1 \ \mu m$ (Fig. 6c). After deposition of the LPCVD second ONO layer (ONO-2), an array of slits, $2 \times 10 \ \mu m^2$, are patterned using reactive ion etching (RIE); these are intended to provide XeF$_2$ access in the next step (Fig. 6d). Then, the sacrificial polySi layers are etched away by XeF$_2$ dry gas; the bulk silicon is partially removed in this step (Fig. 6e). The next step is to seal the access holes with the third ONO layer (ONO-3) that is thickness-controlled to have a mild tensile residual stress of $\approx 42 \ \text{MPa}$ to avoid a buckling of the suspended membrane due to compressive residual stress. The total thickness of sacrificial polySi layers for the narrow channel in the KLP is $\approx 0.8 \ \mu m$, lower than the design value of 1 $\ \mu m$ in Section II. The fabricated chip has a footprint of $12 \times 15 \ \text{mm}^2$. The KHP, KLP, individual stages, IN/OUT ports, and Pirani gauges are shown in Fig. 7. The fabrication yield is typically 60–80% with the available tool set.

During fabrication, the main differences between the KHP and the KLP regions of the chip (Fig. 8) are: 1) the areal density of XeF$_2$ access holes is eight times larger in the KLP than that in the KHP; and 2) the narrow channels are formed by sacrificially etching the thin polySi layer for the KHP and the thin and thick polySi layers for the KLP. The height and the lateral undercut of the wide channel are controlled by the areal density of the XeF$_2$ access holes and etching time. Hence, the wide channel in the KLP is etched deeper than that in the KHP, and the XeF$_2$ lateral undercut profile in the KLP is wider than that in the KHP.

The cross sections of the narrow and wide channels in the KHP and the KLP are examined with scanning electron microscope (SEM) images (Fig. 9). The difference between the cross section of narrow channel in the KHP and KLP is evident from Fig. 9. The KHP uses a narrow channel that is 0.1 $\mu m$ in height and 22 $\mu m$ wide, whereas the KLP uses this further enhanced by an opening that is nominally 0.8 $\mu m$ in height and 12 $\mu m$ wide. In the wide channel (Fig. 9c and d),

### Table III

<table>
<thead>
<tr>
<th>Layer</th>
<th>Thickness ($\mu m$)</th>
<th>Residual stress (MPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPCVD ONO-1</td>
<td>0.5/0.2/0.5</td>
<td>+42.5</td>
</tr>
<tr>
<td>LPCVD ONO-2</td>
<td>0.5/0.2/0.5</td>
<td>+42.5</td>
</tr>
<tr>
<td>PECVD ONO-3</td>
<td>0.7/0.5/0.7</td>
<td>+42.1</td>
</tr>
<tr>
<td>Thin/thick polySi</td>
<td>0.1/0.7</td>
<td>-</td>
</tr>
<tr>
<td>ALD Al$_2$O$_3$</td>
<td>0.2</td>
<td>+304</td>
</tr>
<tr>
<td>Ti/Pt</td>
<td>0.025/0.1</td>
<td>-</td>
</tr>
</tbody>
</table>

Fig. 7. Photograph of an as-fabricated chip. The upper right inset shows a stage, the lower right inset a Pirani gauge and the lower left inset the IN port.
the eight times larger areal density of XeF₂ access holes in
the KLP produces a height of 108 μm compared to that of
36 μm in the KHP. The larger density also produces a wider
channel width of 422 μm in the KLP than that of 358 μm in
the KHP, evident as the silhouette of the undercut in the top
view (Fig. 8); the defined width of the cavity opening in the
photomask (Fig. 6a) is only 300 μm in both KHP and KLP.

IV. TEST RESULTS
A. Methods
The fabricated chip is evaluated in a test chamber to
permit the control of ambient pressure. The chip is wire-
bonded for electrical connections from measurement tools
to Ti/Pt metal pads. The substrate of the Knudsen pump is
attached to a heat sink of the test chamber for thermally
grounding to the room temperature in laboratory environment.
The operating medium is laboratory air. The input port of
the Knudsen pump is sealed, while the output port is open
and vented to the test chamber. The pump is operated by
providing a constant voltage to the Ti/Pt heater; the total
input power to 162 stages is 390 mW and the input power
in a stage is ≈2.4 mW. The temperature increase of the
heater is calculated by dividing the fractional change in
resistance by the thermal coefficient of resistance, α. The
input power levels, as indicated in test results (Figs. 10–13),
are for the heated resistances at equilibrium pressures.

For measuring the vacuum levels at evacuated stages, a
constant current of 4 mA is provided to each Pirani gauge
Fig. 11. Typical test results for the operation of KLP by itself. (a) Equilibrium pressures at various ambient pressures. (b) Transient responses of only P162 at each ambient pressure.

Fig. 12. Variation in input power for the combined operation of KHP and KLP at 760 Torr ambient.

Fig. 13. Reliability test for 37 days of continuous operation. The subscript \( n \) denotes the elapsed days. Each pressure difference from Pirani gauge P1 to Pirani gauges P54, P99, and P162 is normalized to that at the beginning, 0.

and the measured fractional change in resistance is correlated to pressure. The input power level for the heated resistance is \( \approx 2.5 \) mW, which is modest relative to 390 mW for the Knudsen pump heater.

The Pirani gauges are calibrated by a special method, named dynamic calibration, which accommodates the reduced interior pressure and the ambient exterior pressure of the Pirani gauge during pump operation; this is described in [31] in detail. Here, the method is briefly illustrated for P162 and P1; P54, and P99 follow the case for P162. The method uses the response of P162 and P1 under rapid modulation of ambient pressure. 1) P162 provides a response that represents the combined effect of modulated exterior ambient pressure and nearly unchanged interior pressure; this is due to the relative slowness in pressure variation in the interior, as compared to the rapid modulation of the ambient pressure. 2) P1 provides a response that always represents the rapidly modulated ambient pressure because the interior, together with exterior, of P1 is directly exposed to ambient pressure. The multistep-modulation of the ambient pressure yields an equilibrium pressure for which the calibrated responses of P162 and P1 are equal. This equilibrium ambient pressure is exactly equal to the interior pressure of P162. Process-induced variations in responses of P162 and P1 are fitted by a linear regression model. In evaluating the pump performance, the measured pressure values are indicated with error bars for \( \pm 2\sigma \) (where \( \sigma \) is the standard deviation), obtained by error analyses of the residual non-linearity in the regression model, together with repeatability of Pirani gauge sensors. The error bars for \( \pm 2\sigma \) represent a 95.4% confidence interval, assuming normal (Gaussian) distribution [32].

B. Pumping Results

Two sets of tests are performed to evaluate the fabricated devices. The first test evaluates the combined operation of the KHP and KLP. For this test the inlet at stage 162 remains sealed, while stage 1 serves as the outlet. The second test evaluates the KLP by itself. For this test, stage 162 remains as the sealed inlet, while stage 54 serves as the outlet. In order to facilitate this, the supporting membrane at P54 is broken to allow it to vent to the ambient. Therefore, this is a destructive test.

The unheated resistances are \( \approx 126 \) \( \Omega \) for the Knudsen pump heater and \( \approx 140 \) \( \Omega \) for the Pirani gauge. The experimental value of \( \alpha \) is 2,314 ppm/K for the Ti/Pt material used in both heaters and gauges. The temperature increase of the Pirani gauge from the typical fractional increase in resistance of 0.12 at 0.1 Torr is \( \approx 52 \) K.

For the first set of tests (Fig. 10a), 0.39 W was applied to the heater at atmospheric ambient pressure. The temperature
increase of the heater was typically \( \approx 56 \) K; this increase is twice that of the theoretical estimate using FEA (Fig. 5). The pressure at P162 was 0.9 Torr, which corresponds to a total compression ratio of \( \approx 844 \). The equilibrium pressures at intermediate stages were 258 Torr for P54 and 7.8 Torr for P99.

For the second set of tests (Fig. 11a), an input power of 0.29 W was applied. The temperature increase of the heater was typically \( \approx 68 \) K. The equilibrium pressures of P162 were measured at various ambient pressures to mimic various downstream conditions for the KLP. At atmospheric pressure, the equilibrium pressure at P162 was 48 Torr. At ambient pressures from 200 Torr to 3 Torr, the equilibrium pressures ranged from 3.4 Torr to 1.5 Torr, respectively. At an ambient pressure of 0.95 Torr, the pressure reduced to 0.6 Torr. The pre-equilibrium transient behavior was also evaluated during the two sets of tests. Using a Labview™ program, the transient response was automatically recorded every 30 seconds after applying the input power. For the first set of tests at atmospheric ambient pressure, P162 took \( \approx 25 \) hours to achieve equilibrium (Fig. 10b). P54 typically took \( \approx 50 \) hours, i.e., twice as long. Once the heater was turned off, the pressure at P162 returned to the ambient value in \( \approx 10 \) hours, 5 times shorter than the evacuation time of \( \approx 50 \) hours for P54. The transient response for the second set of tests was measured at ambient pressures of 760 Torr, 200 Torr, 60 Torr, and 20 Torr (Fig. 11b). To achieve equilibrium, P162 took \( \approx 5 \) hours, \( \approx 2 \) hours, \( \approx 1 \) hour, and \( \approx 0.5 \) hour, respectively.

The impact of varying input power was investigated using the first set of tests at atmospheric ambient pressure (Fig. 12). The input power was increased to 0.41 W, 0.69 W, and 1.08 W. The temperature increases of the heater went up to \( \approx 68 \) K, \( \approx 120 \) K, and \( \approx 174 \) K, respectively. The equilibrium pressures for P54 were 96.8 Torr, 9.3 Torr, and 4.7 Torr, respectively. The pressures for P162 were 0.90 Torr, 0.80 Torr, and 0.84 Torr, respectively. The highest compression ratio was approximately 760/0.80, i.e., 950, for 0.69 W. For the higher power of 1.08 W, the compression ratio was lower.

**C. Reliability Tests**

To investigate the reliability of a micromachined Knudsen pump, one sample of the full 162 stage Knudsen pump was continuously operated for 37 days, at an input power of 0.46 W. Figure 13 shows that the evacuation levels for each of Pirani gauges P54, P99, and P162 remained within 1% of the value on the first day. Since mechanical failure is unlikely because of the absence of moving parts, the primary failure mechanism could stem from the heating of the Knudsen pump. However, the thermal degradation of the various layers is expected to be very modest because the temperature increases by only 71 K, i.e., to about 96 °C, at 0.46 W. In contrast, the deposition temperature of the various thin films used in the fabrication were 910 °C for LPCVD oxide, 802 °C for LPCVD nitride, 380 °C for PECVD oxide, 380 °C for PECVD nitride, and 250 °C for ALD \( \text{Al}_2\text{O}_3 \). In case of Ti/Pt metal, annealing at 600 °C for 5 hours stabilizes electrical properties [33]. Hence, the Knudsen pump, without any substantial failure mechanism, can provide high reliability and long term operation.

**V. DISCUSSION AND CONCLUSION**

The compression ratio and device volume are benchmarked with other previously reported pumps [4], [7], [12]–[14], [20], [34]–[43]. Using an input power of 0.39 W, the compression ratio of \( \approx 844 \) (= 760/0.9) exceeds the highest previous compression ratio [7], by a factor of \( \approx 17 \), with 3.5 times better power efficiency (Fig. 14). This performance over the other pumps is enabled by: 1) the monolithic integration of 162 stages into one silicon chip, using silicon-based micromachining processes; 2) the two-part design, separately customizing stages into the KHP and the KLP; and 3) the heater design obtained by DOE and FEA.

The effect of \( \Delta P_r \) at high vacuum was evident when using varying levels in input power (Fig. 12). For P162, the measured pressure remained between 0.9 Torr and 0.8 Torr without noticeable reduction, when the \( \Delta T \) increased from 68 K to 174 K. In contrast, for a similar \( \Delta T \) of 50 K to 150 K, the theoretically calculated pressure reduced from 0.39 Torr to 0.04 Torr, which is a larger ratio than that of the measurement. The measured results suggest that the unintended \( \Delta P_r \) in the KLP wide channel fully neutralizes the \( \Delta P_r \) in the narrow channel at pressures less than \( \approx 1 \) Torr. For the theoretical calculation of \( Q_T \) and \( Q_P \), it was assumed that \( T_{Hot} \) gradually decreases to \( T_{Cold} \) along the wide channel (Fig. 1). However, in the pressure range of 1 Torr to 0.1 Torr,
the corresponding mean free path from 49 μm to 511 μm (Table I) encompasses the wide channel length of 250 μm. Therefore, air molecules could be transported from the hot to cold end without collision, resulting in hot molecules entering the cold end of the next narrow channel, and vice versa. This collisionless transport, which could break from the assumed gradual temperature change in the wide channel, could explain why the performance was less than predicted using theoretical calculations. For this reason, a more advanced level of design and experimentation at these low pressures is necessary for further improvement of compression.

The ΔP at high vacuum was also evident from the pressure response of P54. Using similar input powers of 0.39 W and 0.41 W (Figs. 10a, 12), the pressures of 258.4 Torr and 96.8 Torr at P54, respectively, were reduced to 0.9 Torr at P162. The different pressures at P54 might be caused by sample-to-sample variations in channel dimensions. Figure 11 also shows similar P162 pressures of 3.4 Torr and 3.2 Torr when the initial pressures are different as 200 Torr and 60 Torr, respectively. It appears that the pressure at P54 bottoms near 258.4 Torr and 96.8 Torr, also shows similar P162 pressures of 3.4 Torr and 3.2 Torr at P162. The different pressures at P54 might be caused by a mechanism for motionless gas pumps.

Subsequent analyses suggest that at sub-Torr pressure the pump was reliable, with sustained operation over 37 days. Therefore, air molecules could be transported from the hot to cold end without collision, resulting in hot molecules entering the wide channel, could explain why the performance was less than predicted using theoretical calculations. For this reason, a more advanced level of design and experimentation at these low pressures is necessary for further improvement of compression.

In summary, a two-part (KHP and KLP) architecture has been investigated for the 162-stage Knudsen pump. A five-mask single-wafer process was used to fabricate the pump in a small footprint of 12 × 15 mm². Most notably, the high compression ratio of ≈844 was achieved at atmospheric ambient pressure; the input power was limited to ≈0.39 W; and the pump was reliable, with sustained operation over 37 days. Subsequent analyses suggest that at sub-Torr pressure ΔP in the wide channel neutralizes further evacuation. At the cost of additional masking steps in the fabrication process, the architecture may be extended to more than two partitions to improve pumping rate and compression ratio.

ACKNOWLEDGMENT

Facilities used for this research include the Lurie Nanofabrication Facility (LNF) operated by the Solid-State Electronics Laboratory (SSEL) and the University of Michigan.

REFERENCES


Yogesh B. Gianchandani (M’85–SM’05–F’10) is a Professor at the University of Michigan, Ann Arbor, MI, USA, with a primary appointment in the Electrical Engineering and Computer Science Department and a courtesy appointment in the Mechanical Engineering Department. He also serves as the Director for the Center for Wireless Integrated MicroSensing and Systems (WIMS²). Dr. Gianchandani’s research interests include all aspects of design, fabrication, and packaging of micromachined sensors and actuators (http://www.eecs.umich.edu/~yogesh/). He has published about 300 papers in journals and conferences, and has about 35 U.S. patents issued or pending. He was a Chief Co-Editor of Comprehensive Microsystems: Fundamentals, Technology, and Applications, published in 2008. From 2007 to 2009, he also served at the National Science Foundation, as the Program Director for Micro and Nano Systems within the Electrical, Communication, and Cyber Systems Division (ECCS). Dr. Gianchandani is a Fellow of the IEEE.