An atomic layer deposition reactor with dose quantification for precursor adsorption and reactivity studies

T. J. Larrabee, T. E. Mallouk, and D. L. Allara

Additional information on Rev. Sci. Instrum.

An atomic layer deposition reactor with dose quantification for precursor adsorption and reactivity studies

T. J. Larrabee, T. E. Mallouk, and D. L. Allara
Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

An atomic layer deposition reactor has been constructed with quantitative, precision dose control for studying precursor adsorption characteristics and to relate dose quantity and exposure dynamics to fluid flow in both the viscous and molecular flow regimes. A fixed volume of gas, held at a controlled temperature and measured pressure, is dosed into the reaction chamber by computer-controlled pneumatic valves. Dual *in situ* quartz crystal microbalances provide *parallel mass measurement* onto two differently coated substrates, which allows adsorption coverage and relative sticking coefficients to be determined. Gas composition in the reaction chamber was analyzed *in situ* by a quadrupole mass spectrometer. Absolute reactant exposure is unambiguously calculated from the impingement flux, and is related to dose, surface area, and growth rates. A range of control over the dose amount is demonstrated and consequences for film growth control are demonstrated and proposed. © *2013 American Institute of Physics.* [http://dx.doi.org/10.1063/1.4774042]

I. INTRODUCTION

Atomic layer deposition (ALD) is a chemical vapor deposition technique in which alternating pulses of different vapor-phase reactants chemisorb on a surface, growing films in a layer-by-layer fashion with molecular scale precision. ALD has proven to be a unique and successful method because one obtains very precise and reproducible thickness control by simply varying the number of deposition cycles, because the films typically are nearly atomically smooth and pinhole free, and because the films can conformally coat step edges and complex 3D substrates with deep aspect ratios.[1–3] Furthermore, a wide range of materials may be deposited by ALD,[2] mostly performed at moderate temperature, and some materials may even be deposited selectively in certain regions of chemically patterned substrates, known as area-selective ALD.

In an ideal ALD process, a pair of gas-phase precursors reacts separately and exclusively at surfaces. Each precursor must adsorb in a self-limiting fashion, up to a monolayer coverage, after which it no longer deposits, which is often called saturation. The adsorption of one precursor creates complementary surface chemical moieties for reaction with the other, and the different precursors do not meet nor react in the gas phase. This sequential separation of the precursors is most commonly achieved by flushing the residual reactant vapor away with an inert gas purge, or in some cases by thorough evacuation between exposures. Chemisorption should be rapid (compared with the durations of the dose pulses) and complete (for saturative doses). Nucleation of the growing film should be homogeneous everywhere. For an ideal ALD process, adsorbed precursors do not desorb, decompose, or form multilayers that are difficult to remove at moderate deposition temperatures nor do they produce byproducts that etch or contaminate the films. Meeting these requirements means that the exact precursor dose is not particularly important for film growth as long as suf-

ficient precursor dose is used to saturate the surface in each step and sufficient purge or evacuation is provided to prevent gas-phase reaction. Some ALD processes are believed to be close to ideal, such as reaction of trimethylaluminum (TMA) with water to form aluminum oxide (Al_2O_3).[2] However, many processes are far from ideal, such as the growth of metallic Pt from methylcyclopentadienyl(trimethyl)platinum(IV) and O_2 gas,[4] and currently few means exist to quantitatively compare different ALD reactions other than that they are known to work under specific conditions, and not work otherwise.

Reactor geometries, precursor and inert gas flow rates, and valve configurations vary considerably among different reactor designs. In some designs, solid or liquid precursors are placed in an open boat in a heated tube;[5,6] in others a bubbler is used similar to continuous-flow CVD systems;[7] and in perhaps the simplest configuration, known as "vapor draw," an evacuated, closed source vessel has a single valve that opens to flow into the deposition area, where flow occurs via a difference in pressure from the (higher) vapor pressure of the precursor to the (lower) reactor base pressure.[8–10] Despite these differences, the variable commonly used to describe the quantity of vapor in a dose is the duration (in seconds) of the open valve(s) allowing precursor flow into the deposition area. This convention does not lead to a quantitative description of the number of moles of vapor in a dose, nor does it say anything about the actual precursor exposure resulting from a dose. For an ideal ALD process at a fixed temperature and with a given pair of precursors, the pulse times are typically determined empirically for optimum growth on a planar substrate for a particular reactor, and for this reason quantitative dose amounts are typically not specified. However, in non-quantitative systems, several drawbacks exist. For one, little can be determined about the efficiency of reactant use. Secondly, predictive capability is problematic for relating carrier-gas flow rates, pulse durations, purge times, and temperatures to reactant exposures needed to obtain uniform, conformal ALD-growth. Thirdly, in practice, it is difficult to

relate the empirically derived conditions reported in one reactor to the next. Finally, difficulties may be encountered in determining growth conditions for several more challenging cases. Such cases include very low temperature depositions; depositions onto high-aspect ratio, high-surface area, or large area substrates; or depositions using reactants that deviate from an ideal ALD growth. Recently, the importance of dose control was demonstrated even in the example of what was thought to be a highly ideal ALD reaction, the deposition of Al_2O_3, from TMA/water, which has been reported to be not perfectly saturative at 125 °C.[11]

A better understanding of detailed dose and exposure information would also be particularly valuable for area-selective ALD, or additive pattern transfer, whereby differences in the sticking probabilities on two different surface chemical groups allow selective growth, or in other cases where nucleation differences exist from one substrate to the next, i.e., "substrate inhibited growth" or "substrate enhanced growth." Such quantification would also be useful for applying ALD to the controlled deposition of nanoparticles,[12] and structured films, as in catalysis applications,[13] nanolaminates,[14] or enhancing growth on particular crystal faces[15,16] for anisotropic nanostructures, precise dopant control, etc.

Toward these ends, we have designed, developed, and constructed an ALD reactor with a unique quantitative dosing system, where absolute number of moles of gas per dose are both measured and controlled. This new system is capable of (1) systematic study of precursor reactivity with the growing film, and initial reactivity with particular substrates, (2) mass balances on reactants, (3) more thorough understanding of how to design to maximize precursor utilization and minimize purge times, (4) measurement of the sticking probability of precursors (when coupled to an *in situ* measurement of adsorption), (5) real-time monitoring of the reproducibility of dose quantity as precursor sources are depleted or other conditions change, and (6) applications of intentionally dosing to less than a saturating dose in a controlled fashion, some of which are proposed below (in Sec. VI). Critical to this design is that the dose quantity can be varied reproducibly regardless of reactor base pressure, gas flow rate, and that the same dose amount can be used for viscous flow and high vacuum operation. We demonstrate these capabilities by a series of tests with TMA/H_2O deposition cycles for Al_2O_3 films.

II. THEORY OF MEASUREMENTS

A. Quartz crystal microbalance (QCM) and adsorption

The QCM has been a particularly valuable *in situ* diagnostic in ALD systems for real-time monitoring of adsorption/desorption phenomena and providing film-thickness determination. The interpretation of the mass gain measured during a half-cycle of ALD growth has proven to be more subtle than in ballistic, high-vacuum deposition systems due to several factors; the high temperatures involved, the influence of small temperature changes on the resonance frequency, control over the exact area of the deposition, and the high precision and relatively fast dynamics of measurement

required. Most of these challenges, however, have been overcome with simple, but often elegant solutions,[8,17,18] including gluing the active face of the crystals to seal the area of deposition, backside purging the sensor housing with inert gas, and using polished crystals to minimize the difference between the geometric and molecular surface area. The use of $GaPO_4$ piezoelectric resonators in place of quartz has greatly improved high-temperature stability.[19] In the present work it was found, particularly at high vacuum, that it was necessary to attach an internal heater and temperature controller to achieve and maintain the temperature of the crystals at the temperature of the substrate under all conditions. For thin, rigid, ALD-deposited films, the Sauerbrey equation[20] is a valid relationship for frequency change, ΔF, and change in adsorbed mass on the active area of the crystal, Δm,

$$\Delta F = \frac{-2NF_0^2}{A\sqrt{\rho_q \mu_q}}\Delta m, \tag{1}$$

with ρ_q, the density of quartz; μ_q, the shear modulus of quartz; A, the area of the electrode; F_0, the resonance frequency of the fundamental; and N, the oscillator overtone number ($N = 1$ for the fundamental, $N = 3$ for the third overtone etc.).[20–22]

To relate the molecular adsorption, D [molecules/m^2] to the mass measured by the QCM, m_{QCM}, dimensional analysis shows that one needs Avogadro's number, N_A, and M, the molar mass of the adsorbate, which gives

$$\left[\frac{\text{molecules}}{\text{m}^2}\right]\left[\frac{\text{kg}}{\text{mole}}\right]\left[\frac{1 \text{ mole}}{\text{N}_A \text{ molecules}}\right] = \left[\frac{\text{kg}}{\text{m}^2}\right], \tag{2}$$

$$\frac{DM}{N_A} = m_{QCM}. \tag{3}$$

This equation is valid for adsorbates of well-defined molecular mass, such as physisorbed inert gas molecules, but for dissociatively chemisorbed adsorbates the molecular mass is, in general, different than the gas phase species, a case typical in ALD and other reactive depositions. A simple modification to use the average molar mass of the adsorbed species, m_{avg}, generalizes Eq. (3).

$$\frac{Dm_{avg}}{N_A} = m_{QCM}. \tag{4}$$

For systems with saturative adsorptions, m_{avg} may be measured by the maximum mass gain on the QCM, at surface coverage $\theta = 1$, in combination with a measure of the number of molecules per unit area. For measurement of the relative adsorption onto two different types of surfaces, our system has two *in situ* QCMs. When these are at the same temperature and receive identical flux of precursors, this *parallel mass measurement* allows the study of growth differences between the surfaces, including quantities such as a relative sticking probability.

B. Impingement flux and dose

To be general, consider a gas flux containing a mixture of components. The quantitative amount of a particular component that a substrate receives from a dose is related to the

molecular impingement flux of the component, as defined by the Knudsen equation from kinetic theory using the ideal gas law,

$$J_i = \frac{p_i N_A}{\sqrt{2\pi M_i R T}}, \qquad (5)$$

where J_i is the impingement flux of the ith component in the dose mixture, in molecules m^{-2} s^{-1}, p_i is the partial pressure of the ith component, M_i is the molar mass [kg mol^{-1}], and R and T are the gas constant and absolute temperature, respectively.[23,24] As an instantaneous quantity, J_i must be integrated over time to obtain the *net* exposure, Q_i, in molecules m^{-2}, resulting from a dose.

$$Q_i = \int_{t_1}^{t_2} \frac{p_i N_A}{\sqrt{2\pi M_i R T}} dt. \qquad (6)$$

If p_i is known as a function of time, at constant temperature this simplifies to

$$Q_i = \frac{N_A}{\sqrt{2\pi M_i R T}} \int_{t_1}^{t_2} p_i dt. \qquad (7)$$

While a correlation must exist between the dose in moles and the *net* exposure resulting from that dose in moles/m^2, this relationship is system specific, depending on the geometry, flow conditions, and pumping rate, but it has generally not been presented in literature descriptions of reactors. This information is presented herein for this particular reactor design (Sec. IV B).

The *net* exposure resulting from a dose also does not describe the *dynamics* of that exposure, i.e., the residence time for a dose and its relationship to the size of the dose, or the carrier gas flow rate, the pumping rate, etc. Many aspects of reactor design and operating conditions affect the dynamics of exposure, and these dynamics are not simply related to the duration of a valve open time, as is usually presented as the only measure of dose duration. A useful methodology for further measurement of the duration of the majority of the dose is presented for this reactor design, taking advantage of the time dependence of the adsorption and exposure measurements (in Sec. IV C).

III. DESCRIPTION OF REACTOR AND DESIGN CHARACTERISTICS

A. Dosing section

The reactor consists of two independently controlled gas manifolds: a metal precursor source doser (bottom of Figure 1) and a reactant source doser (top of Figure 1). The latter is typically charged with water, but also could be used to dose oxygen, hydrogen, or any complementary gas or vapor reactant). Each gas manifold has a mass flow controller for carrier/purge inert gas. Each precursor/reactant source is connected to a dose bottle, volume 1 or volume 2 in Figure 1. Capacitance manometers monitor the pressure in each dose bottle; a 10 Torr full-scale gauge (internally temperature controlled at 100 °C) for the metal source bottle and a 100 Torr full-scale gauge (internally temperature controlled at 45 °C) for the water bottle. Each dose bottle has a dose valve (valves V1 and V7) that feeds into the reactor through heated tubing.
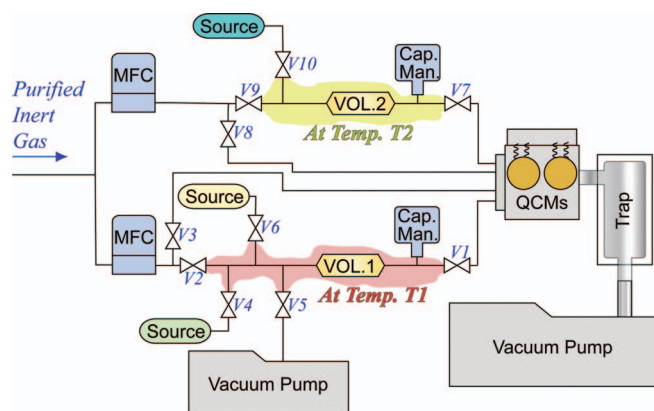

FIG. 1. Reactor overview with detail of quantified dosing section. See text and Table I for description of fill, measurement, and dose and purge functions. Dose quantification and control occurs in real time. All valves in this figure are rapid action, with timing controlled by computer. High vacuum conductance liquid nitrogen-filled trap prevents ALD reactants from entering pump and oil from backstreaming into reactor.

The total volume of precursor gas comes from each complete dose vessel, consisting of bottle, tubing, connecting valves, and gauge connection; which is held at a selected temperature by use of thermocouples, heating tape, and a PID (proportional integral differential) temperature controller. In this way, the dosing manifold forms a complete dose vessel with fixed, calibrated volume, controlled temperature, and measured pressure. The manifolds and valves are mechanically mounted on a sliding rail, so that the bottle may be changed to insert different volume bottles.

B. Operation

Typical operation of the dose manifold consists of sequential fill, dose, purge, and evacuation steps (the latter may be concurrent with the purge step). Table I indicates the states of the valves (Figure 1) for these different steps for the metal precursor half of the ALD process. A similar procedure is applicable for the co-reactant (water) step.

During the dose step, the dose vessel is flushed with inert gas so that total quantity of reactant is rapidly dosed into the reactor. Separate pumping of the metals dose vessel during the evacuation step allows very small amounts of vapor, drawn from the source at pressures less than the reactor base pressure, to be introduced to the inert gas stream, and still be rapidly entrained.

C. Design advantages

Advantages of this design include (i) that the dose quantity can be varied reliably and reproducibly for any system

TABLE I. Typical viscous flow operation of the reactor.

Operation	Valve settings
Fill	V6, V3, V8 open—others closed
Dose	V2, V1, V8 open—others closed
Purge	V3, V8 open—others closed
Evacuation[a]	V3, V8, V5 open—others closed

[a]Optional for filling to pressures less than base pressure.

base pressure, gas flow rate, and in the same fashion for viscous flow and high vacuum operation, (ii) that the dose vessel can provide an *in situ* measurement of the precursor vapor pressure, (iii) that neither the amount of the dose nor the rate of entrainment into the gas stream depends on the evaporation rate of the precursor or the pressure difference from the precursor's vapor pressure to the reactor base pressure, and (iv) that one precursor may be allowed to fill while the other is dosing or purging, which is particularly useful for low-volatility and slowly evaporating precursors. An more subtle aspect of this design is that the separation into two manifolds means valves that fill and dose reactants are not exposed to both reactants, but instead that the metals manifold valves see only metal precursor and inert gas, and the water manifold valves see only water and inert gas; hence the deposition process will not occur on the internal valve surfaces. Condensable vapor streams are exposed only to heated valves or tubing before entering the reactor. Not shown in Figure 1 is also a separate line, bypassing the reactor, connected through a manual valve directly to the pump. This is for the purpose of purging and evacuating each dose vessel on startup and allows removal and installation of sources without interruption of the reaction chamber.

D. Reaction chamber and pumping section

Further detail of the reaction chamber and pumping system is in Figure 2. The two gas manifolds feed into a laminar-flow mixing/preheating stage which consists of a 13.5 in. (34.3 cm) long, 2.37 in. (6.02 cm) ID stainless steel tube with a 4-element, 13.375 in. (34.0 cm) Kenics^TM static mixer inside, and 4.5 in. conflat (CF) vacuum flanges. Three 900 W, 240 V band heaters are wired in parallel and used with two

thermocouples to monitor and control the temperature of this stage with a PID controller, and it is insulated with Zetex^TM ceramic fiber tape. This stage ensures that the gas stream has uniform composition, uniform temperature up to the reaction temperature, and uniform flow velocity profile at its exit. Temperature uniformity of the gas flow impinging on the QCM has been shown to be important to prevent convection from the gas dose from causing artificial apparent mass changes on the QCM.[17] The reaction chamber consists of a cross-flow design made from a 4.5-in. CF cube. The reaction chamber is warm-walled, heated by external flange heaters, heating tapes, and conduction from the substrate heater, and insulated by Zoltec Pyron^TM carbon-fiber based blankets (of the type used in welding applications). The substrate holder/heater is a stainless steel cylinder welded to a 4.5-in. CF flange with five 100 W cartridge heaters connected externally to the vacuum, and two thermocouples inserted to within 3 mm of the surface at the leading edge and center. The center thermocouple is used for PID control of the heaters, and the edge-to-center temperature uniformity is better than 2 °C at all temperatures. It is mounted to hold a 2 in. wafer upright in nearly the center of the cube. Samples and mass sensors are loaded from a 4.5-in. CF door with a Kalrez^TM 7075 high-temperature O-ring, which is the only non-metal vacuum seal on the system. The reaction chamber has two mutually exclusive exhausts, one for viscous laminar flow to a (nominal 15 m^3/h) rotary-vane vacuum pump, the other to a (nominal 210 l/s) turbomolecular pump for UHV operation. High vacuum operation then consists of gate valve GV1 (Figure 2) being open to Turbo 1, and the vacuum valve to the mechanical pump closed; and vice versa for viscous operation. A second turbopump, Turbo 2, is used to differentially pump the quadrupole mass spectrometer (QMS). A second rotary-vane pump (nominal 21 m^3/h) is used to back both turbopumps, and to evacuate the dose vessel (in viscous flow mode). Two capacitance manometers, which are internally temperature controlled at 45 °C, are connected to the reaction chamber (not shown in Figure 2), a 100 Torr full-scale and 1.0 Torr full-scale. A 20 mTorr full-scale capacitance manometer (depicted in Figure 2) is connected on the turbopump-side of the high-vacuum gate valve, GV1. The 20 mTorr gauge and a Bayard-Alpert ion gauge (IG1) are thus used only in high vacuum mode, and they are behind an additional pneumatic 2.75-in. gate valve, GV2, which for some experiments is dynamically closed during one reactant dose to prevent the deposition from occurring on the gauges. Not shown in Figure 2, but also connected to the reaction chamber is a UHV variable leak valve behind a 2.75-in. CF gate valve, which can be used for precision leaks for isotherm experiments and continuous trace reactant introduction. The high vacuum gauges were calibrated by flowing N$_2$ into the reactor through the leak valve and the mass flow controllers. The true reactor pressure in high vacuum mode thus comes from a calibration curve established by temporarily placing the 20 mTorr capacitance manometer on the main chamber, and is slightly higher than the raw reading on either the 20 mTorr gauge (in its normal position) or the reading from IG1. IG1 was corrected for N$_2$ sensitivity using overlapping pressures with the capacitance manometer (from $10^{-5} < P < 10^{-3}$ Torr).
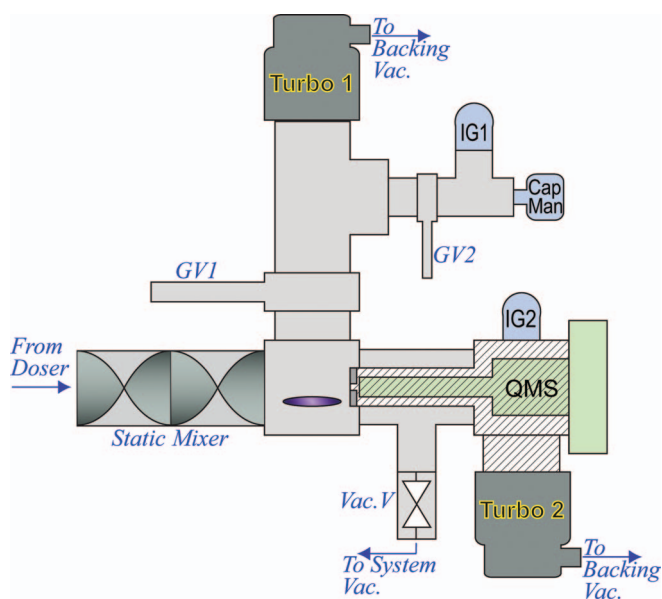

FIG. 2. Reaction chamber and pumping section detail. Gate valve 1 and the system vacuum valve are for high vacuum and viscous flow type operation, respectively. The quadrupole mass spectrometer (QMS) is differentially pumped by turbopump 2 to a pressure of $\sim10^{-8}$ Torr, as depicted with a crosshatch pattern, through a removable internal vacuum flange with a pinhole. Not shown are the QCMs, capacitance manometers close to the substrate, and precision leak valve.

E. Diagnostics and computer control

The QMS is a Hiden triple mass filter model 3F, with mass range of 1–300 amu, equipped with an electron multiplier detector. During typical operation at a system pressure of ~1 Torr, it is differentially pumped through a 30 μm orifice, for an operating pressure of about 2.8×10^{-7} Torr measured by nude Bayard-Alpert gauge IG2 in Figure 2. The QMS is capable of a minimum detectable partial pressure of ~10^{-15} Torr (inside the differential pumping stage, in this case).

The two QCMs (Maxtek/Inficon BSH-151 bakeable holders) were custom modified by welding them close together, drilling 1/32 in. (0.79 mm) holes inside the drawers from the water cooling lines to provide an inert gas purge behind the back of the crystal and electrodes, and welding them onto a 4.5-in. CF flange. During viscous flow operation, inert gas was delivered through a 100 sccm MFC, typically at 30 sccm. For optimum temperature measurement and control attachment pads for thermocouples were welded less then 3 mm from the crystal faces, and a 572 W, 10 mm × 50 mm × 2.5 mm AlN ceramic heater was attached to the rear of the crystal housing internally. This was controlled with a PID temperature controller with K-type thermocouple to within ±0.2 °C over the range from room temperature to ~320 °C during the course of a deposition with carrier gas, and to better than measurement resolution (0.1 °C) in UHV over several hours. The sensor face of the crystals is parallel to and opposite the substrate holder, at a distance of ~21 mm, so that they experience nearly identical flow and flux. Quartz crystals used were 6 MHz polished AT-cut type with RMS surface roughness of ~5 Å for experiments <~150 °C, and polished Y-11º cut GaPO$_4$ crystals were used for experiments >~150 °C. To calibrate which crystal type was most stable at a given temperature, the series resonance frequency of the fundamental as a function of temperature was measured for each type by detecting minimum transmission impedance using a function generator and RF amplitude-sensing circuit.

Valve dosing was computer-controlled using National Instruments (NI) LabVIEW software and NI data acquisition hardware, which was also used to record all the pressures from the capacitance gauges, and interface with either a Maxtek TM-400 thickness monitor or external oscillators circuits and HP/Agilent 53181A frequency counters for the QCM measurements. The latter, operating at the 3rd overtone of the fundamental (18 MHz), proved more sensitive and offered greater flexibility between measurement accuracy and response time. Frequencies were converted to masses using Eq. (1).

IV. DEMONSTRATION OF REACTOR

A. Control of dose

The exact dose vessel volumes include the internal volumes of the bottles, valves, and connections and can only be estimated geometrically. Hence, calibration of the accurate dose volumes was performed by measurement of the pressure rise verses time as the empty vessels were filled at a fixed mass flow rate, using the mass flow controllers. Using the

ideal gas law, and differentiating with respect to time, t, at constant V, T:

$$V \frac{\partial P}{\partial t} = RT \frac{\partial n}{\partial t}. \qquad (8)$$

Controlling the mass flow rate, $\dot{m} = dm/dt$, is the same as controlling the molar flow rate, $\dot{n} = dn/dt$, so with molecular weight, M: $Mn = m$, so $\dot{m} = M\dot{n}$, and we may write:

$$\dot{n} = \frac{\dot{m}}{M}. \qquad (9)$$

Integrating Eq. (8) over time from t_i to t_f:

$$V \int_{t_i}^{t_f} \frac{\partial P}{\partial t} dt = RT \int_{t_i}^{t_f} \frac{dn}{dt} dt \qquad (10)$$

and using Eq. (9), this becomes

$$V(P(t_f) - P(t_i)) = RT \frac{\dot{m}}{M} (t_f - t_i), \qquad (11)$$

which can be rearranged to

$$\frac{M \Delta P}{RT \dot{m} \Delta t} = \frac{1}{V}. \qquad (12)$$

For a constant mass flow, $\Delta P / \Delta t$ is constant. After a short time, the P becomes linear with t and is used to calibrate the volume using Eq. (12). Five different nitrogen fill cycles were recorded at a given mass flow rate, each yielding a volume. The procedure was performed at several different mass flow rates, and these volumes were then averaged. The final calibrated volume of the metal precursor source dose vessel used (corresponding to VOL. 1 in Figure 1) was 73.9 ± 0.5 cm^3, and of the water source dose vessel (VOL. 2 in Figure 1) was 66.7 ± 0.2 cm^3.

Control over the dose amount is achieved by controlling the fill pressure, since the volume and temperature are fixed. For volatile liquid and solid sources, the fill time and source temperature are sufficient to provide control of the fill pressure. Figure 3 demonstrates the control over the dose achieved for TMA vapor for different fill times.
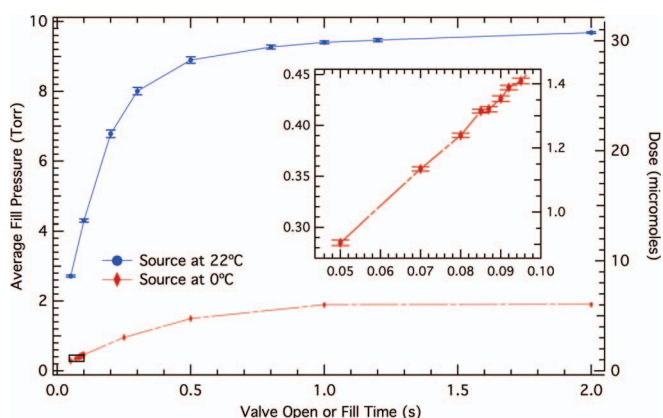

FIG. 3. Control of dose amount achieved from drawing vapor off of a liquid TMA source, with fill pressures (left axis) and dose amounts (right axis) calculated using calibrated dose volume and fixed vessel temperature of 100 °C. Upper curve represents room temperature source and lower curve with source cooled to 0 °C. Points are averages of 30 measurements, error bars are ± 1 standard deviation of the 30 measurements and are primarily to illustrate reproducibility of fill. Inset zoom is of lower left boxed region of lower curve.

For non-condensable gases, the fill time and placement of a restricting needle valve (0.004 C_v flow coefficient) immediately before the fill valve, to slow the fill rate to a time scale that the fill valve timing effects, provide control of the pressure. In this case use with a low pressure regulator (∼10–15 psi (gauge) is typically supplied) on the cylinder is most effective for fill pressures ≪ atmospheric.

B. Measurement of impingement flux

The net exposure resulting from a dose was measured by numerically integrating the impingement flux, as in Eq. (7), with the bounds of integration encompassing a pressure peak resulting from a dose into high vacuum. To capture the most accurate pressure, the capacitance manometer with full-scale range of 20 mTorr was used. It was determined that due to the long, slow decay of the pressure peak, and a slight zero-drift and noise in the measurement, it was difficult to select the t_f of the integration, and objectively measure the integral reproducibly. This was especially true for small doses. Hence, instead, pressure peaks were integrated after fitting to a function, a convolution of exponentials, of the general form:

$$f(t) = \frac{k_1 k_2 (e^{-k_1 t} - e^{-k_2 t})}{k_2 - k_1}. \tag{13}$$

In this function form, for $k_1 > k_2$, k_1 is associated with an exponential rise time and k_2 is an exponential decay. Larger k_1 results in a faster rise and smaller k_2 results in a longer decay.

Figure 4 illustrates the process of quantifying the dose and integrated impingement resulting from the dose.
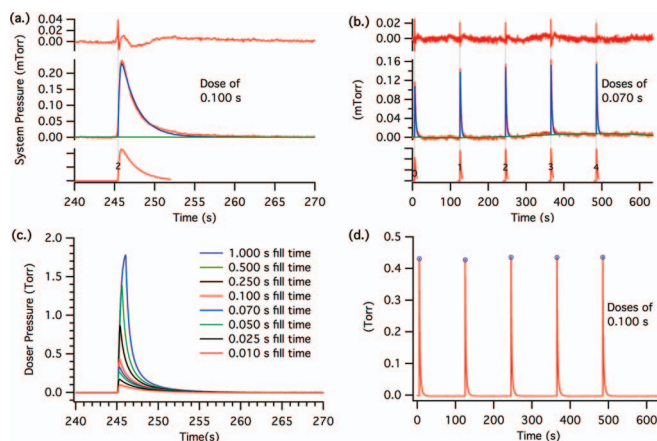

FIG. 4. Example of method for calculation of integrated impingement flux (upper half) and for quantification of dose (lower half). (a) Function fit of a single system pressure peak to form of Eq. (13). Top plot is the residual, middle plot is the data and fit together; bottom graph is the fit function alone. (b) Illustration of the fits of 5 such pressure peaks. (c) Quasi-static doser pressures controlled with fill time. The rise is during the fill, the decay is the emptying of the dose into the reactor. (d) Doser fill consistency example. Circles represent the locations of peak pressure and hence total dose vessel fill pressure because at the peak, all valves enclosing the dose vessel are closed momentarily. Using the fixed volume and temperature of the dose vessel, true absolute doses are converted to micromoles using the ideal gas law. All doses shown are of TMA vapor drawn from a 0 °C source into the dose vessel, held at 100 °C. Fills and empties of both the reactor (top half) and dose vessel (bottom half) are into high vacuum (P < 10^{-5} Torr); negligible compared to the quantities measured.
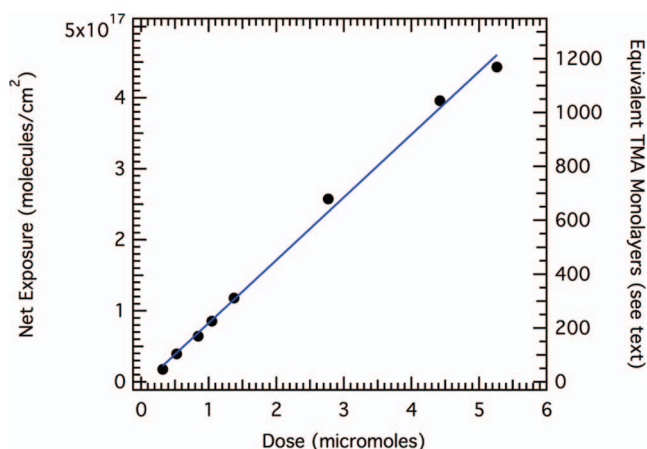

FIG. 5. Relationship between the net exposure, measured as the integrated molecular impingement flux (Eq. (6)) and the quantified dose in micromoles. Each point represents the average of 5 integrated impingement fluxes and 5 dose amounts per fill time, as depicted in Figs. 4(b) and 4(d). The straight line is a least-squares fit to the data. The number of equivalent TMA monolayers that correspond to a given net exposure is plotted on the right axis. This is determined by dividing the net exposure by the estimated monolayer packing density of TMA, 3.79×10^{14} molecules per cm^2 (see text). As an aside: note that the actual number of equivalent TMA monolayers of flux at the QCM and substrate is smaller than this number *during* a deposition because significant surface area exists along the flow path prior to the substrate, and this reactive surface area would consume a portion of the quantified dose before reaching the substrate. This upstream surface area was extensively saturated with TMA immediately prior to the measurements in Figure 4 used in this correlation.

This process developed the relationship between the *net exposure* resulting from a dose to the *dose amount*, in Figure 5, which is linear for the region investigated.

This relationship is important because it allows an accurate estimate of the exposure resulting from a dose by simply knowing the dose amount and because it elucidates what the range of control of the net exposure is in terms of dose, for this particular geometry. The number of equivalent TMA monolayers corresponding to these exposures was calculated by estimating the TMA monolayer packing density, in molecules per cm^2. From the liquid density[25] of TMA, 0.743 g/cm^3, using the approximation that TMA can be represented as well-packed spheres in the liquid, we used the volume fraction of spheres of $\pi/(3\sqrt{2}) \cong 0.74048$ to determine a spherical molecular diameter, which was 6.1078 Å. The monolayer packing density was estimated as ∼3.79 $\times 10^{14}$ molecules/cm^2 on the basis of a 1 molecular-diameter-thick slice of the liquid. The exact monolayer packing density will, in general, be specific to the substrate onto which the monolayer would adsorb, because of differences in numbers of binding sites, mechanism, etc., but this number serves as a geometric estimate.

C. Exposure dynamics

In order to investigate the time scale of reactant exposure as it relates to dose size in different flow regimes, and with different carrier gas flow rates, etc. for comparison to reaction kinetics, we sought to use the *in situ* diagnostics for temporal information, in addition to net exposure.

In the most ideal case of extremely fast adsorption and reaction kinetics, the duration of the exposure and its time trajectory in a flow reactor are of little importance. However, in ALD reactor design and operation, the conflicting goals of minimizing the cycle time and yet maximizing the growth per cycle are always present, and not all reaction chemistries are fast compared to these time scales. Conformally coating deep aspect ratio structures generally requires much longer cycle times, and in cases where mean free paths are comparable to nanostructure dimensions, understanding high vacuum ALD matters. In all reactors, a reactant pulse will broaden in time by diffusion, by gas hold-up in inefficiently purged/evacuated regions, as well as by differences in pumping speed, etc. Many factors affect the efficient utilization of reactant gas, inert gas, and time, such as valve flow coefficients and vacuum conductivities, as recently demonstrated.[26] Relatively few diagnostics have been applied to explicitly look at the time dependence of the gas-phase reactant, as it relates to the surface reactions, with the notable exception of recent IR techniques.[27–29] QMS techniques have been applied to ALD for some time,[30–32] but generally qualitatively to look only at the presence of particular species in the gas phase, for understanding reaction mechanisms, rather than to probe actual reactant exposure dynamics.

One measure providing such dynamical information is the time scale of the exposure of a reactant dose to the substrate, the *true* dose duration. A useful way of estimating it is by the full width at half maximum (FWHM) of the impingement flux peak resulting from dose. Figure 6 illustrates the time evolution of differently sized doses into a high vacuum flow without carrier gas, demonstrating that larger doses have shorter true dose durations under these high vacuum conditions, though net exposure is proportional to dose size, as previously discussed.

For some particular value of the dose size, the duration of exposure and net exposure amount are sufficient for saturative

growth. However, in the case of high vacuum flow (where the Knudsen number, $K_N > 1$), we see from Figure 6 surprisingly as the dose increases, resulting exposure durations measured by the FWHM do not increase with increasing dose size. Using FWHM to demonstrate that even though larger doses integrate to larger exposures, the *duration* of the *largest part* of the exposure is smaller because the exponential decay is approximately the same as for smaller doses. This means that for true dose durations shorter than the kinetics of reactive sticking allows, increased dose is wasted. The physical basis of this counterintuitive effect is a consequence of the vacuum throughput relationship $Q = SP$, where Q is the throughput in Torr l^{-1} s^{-1}, S is the pumping speed in l s^{-1}, and P is the total pressure in Torr.[23]

The combined effect upon reactive adsorption of exposure durations and net exposures may be systematically decoupled with this reactor design, both in high vacuum flow and viscous flow using the true $P_{tot}(t)$ and the $p_i(t)$ from the QMS. In general, true dose durations of reactant exposure to the substrate as measured by this FWHM method are anticipated to have the opposite relationship to dose size than in Figure 6 during viscous flow, as profiles indicated by IR results suggest.[28] In continuum flow of carrier gas through the reactor, one would expect that a *smaller* well-entrained, minimally diffused dose would have a *shorter* true dose duration, however transport phenomena dynamics of the particular design will dictate this relationship.[33] Never the less, in *any flow regime*, if true dose durations to the substrate are shorter than sticking times for doses resulting in net exposures larger than saturative impingement, potentially expensive reactant is wasted downstream of the substrate, so such a measure is useful.

Discrimination of *total exposure* from *duration of exposure,* as it relates to adsorption reaction kinetics, deserves further exploration, such as this FHWM measurement, which offers a meaningful comparison tool. In the example above, ~71% of the total exposure occurs during the time period of the FWHM, so using the FWHM to estimate the *true* dose duration is reasonable.
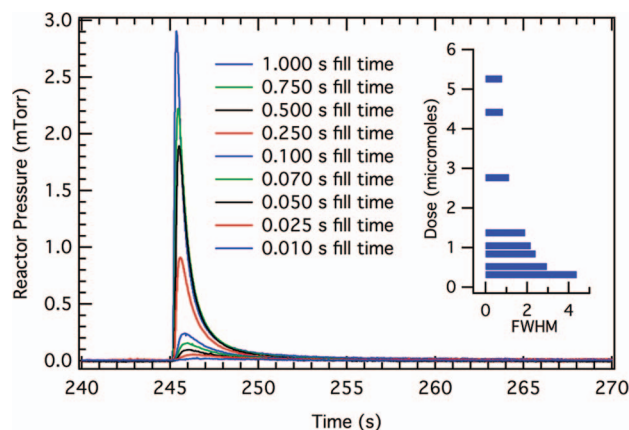

FIG. 6. Dynamics of differently sized doses into high vacuum flow, without carrier gas. Dose size increases with fill time as does peak reactor pressure resulting from the dose. An estimate of the *true* duration of a dose's exposure to the substrate, the full width at half maximum (FHWM) of each pressure peak, is displayed in the inset, demonstrating the inverse relationship between dose size and dose duration. Note this effect will be enhanced for larger ΔP as the dose travels through the reactor and would not necessarily follow the same relationship for continuum flow situations (see text).

D. Calibration and quantification of QMS

Since ALD is most commonly performed in viscous flow, it is desirable to be able to relate actual impingement flux and exposure dynamics of precursors under these conditions. For this, one needs to know the true partial pressure of reactant in the gas stream as a function of time. The QMS used here outputs the counts per second of the electron multiplier detector. To develop a calibration to convert this to partial pressure, the vapor of a reactant was bled through the leak valve into the reactor under steady state conditions with the reactor being pumped by turbopump 1, monitoring the pressure from the 0.020 Torr range capacitance manometer (Figure 2). Figure 7(a) is a mass spectrum obtained under these conditions, with electron impact ionization at 70 eV. Figure 7(b) is an example of how the QMS was calibrated by recording several different pressures and correlating them to the mass spectrometer's counts/s at m/z = 57 amu.
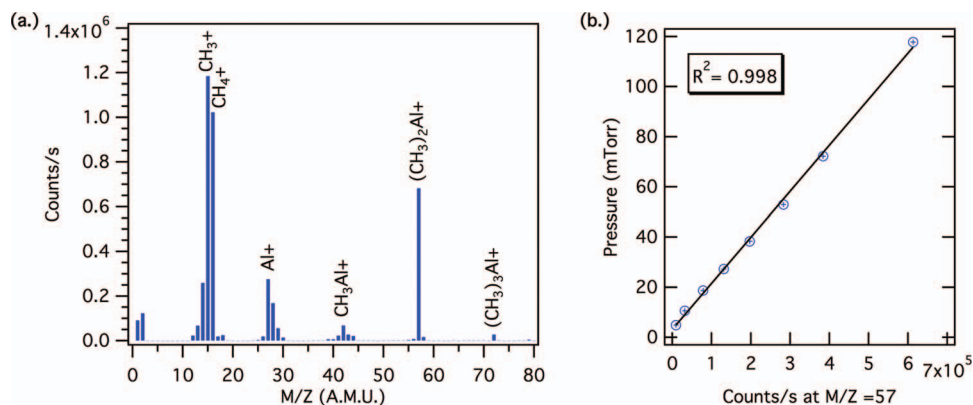
FIG. 7. (a) Mass spectrum obtained for trimethylaluminum (TMA). (b) Calibration curve of actual reactor pressure of pure TMA vapor to QMS signal in counts per second at mass to charge ratio 57, with linear fit. This m/z value is the most abundant Al-containing ion. Larger signals at m/z = 15 (CH_3+) and m/z = 16 (CH_4+) overlap with the CH_4 product of reaction, so would be poor choices for a TMA pressure calibration.

A demonstration of monitoring the ALD reactions in this system for TMA and water at 125 °C in 300 sccm of nitrogen at ~1 Torr is given in Figure 8. This setup allows for an accounting of fate of a measured dose of reactant. Specifically, the *ensemble* of a calibrated QMS, a QCM, and dose quantification and control now allows detailed mechanistic molecular mass balances and their relation to $J_i(t)$ and gaseous byproduct production, even with the inherently non-steady state ALD process.

It was noted that in performing the QMS calibration for the m/z = 15 and 16 species, quantification of the partial pressure becomes more complicated. Because these peaks may be from contributions from the TMA reactant or the surface reaction product in this case, the partial pressure calibration was performed with both TMA alone and pure methane gas alone. In Figure 9, we note that the m/z peaks at 15 and 16 have different relationships to gas concentration depending on whether they arise from TMA or methane.

This would be expected to be a problem for complete quantification of the QMS data for many ALD reactants because the reactant molecule will likely fragment to produce m/z ratio peaks that overlap with the reaction product. Two possible solutions for achieving calibration allowing for required quantification for a molecular mass balance include: (1) use of isotopically labeled co-reagents like D_2O or $H_2^{18}O$ in place of H_2O, yielding m/z ratios for product species different than those arising from, for example, TMA fragmentation and (2) altering the EI energy, because at lower EI energies (such as 20 eV instead of 70 eV), a higher relative portion larger m/z ratio species from the reactant molecules will be present due to less fragmentation.[34] For this reason, the partial pressures in Figure 8 are expressed in arbitrary units, and further work involving the use of isotopically labeled species is underway for true partial pressure analysis.

E. Parallel mass measurement for relative sticking probabilities

This setup is useful for measuring the relative sticking probabilities onto two differently coated substrates, a useful technique for quantifying growth selectivity and designing surface structures and conditions for area-selective ALD growth. As a simple demonstration of such a parallel mass
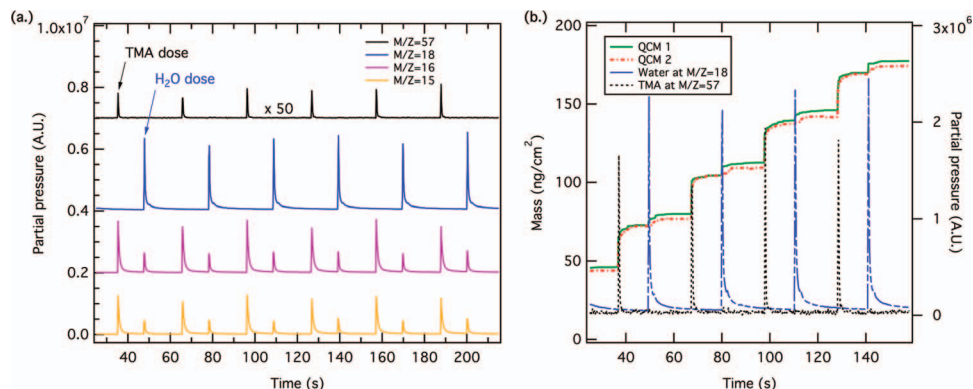

FIG. 8. (a) Monitoring m/z = 57 for TMA, m/z = 18 for water, and m/z = 15 and 16 during the course of 6 typical ALD cycles for Al_2O_3 growth. Because m/z = 15 and 16 are associated with methane, contributions at each peak during a TMA dose may be due to the fragmentation of TMA molecules in the ionizer of the QMS, or the product of reaction with walls of the reactor, or even possibly CVD-like reactions if the water purge step was insufficient. During the water dose, they would most likely be associated with releasing ligands from surface bound TMA molecule fragments. (b) Four typical ALD cycles showing the QCMs masses (left axis) correlated to the QMS peaks from TMA and water (right axis). Total calibration of all of the QMS partial pressures, combined with quantified dosing and the QCMs, allows for accurately accounting of where and when the reactant is consumed, and byproducts of the reaction produced (in progress as of this writing).
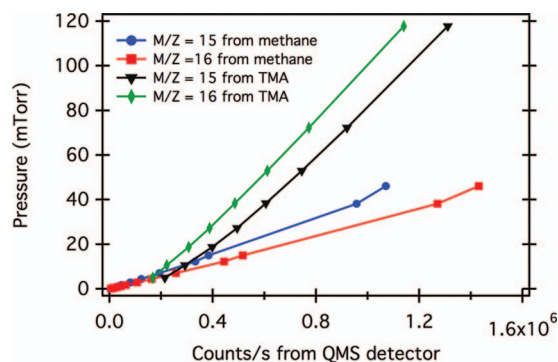
FIG. 9. Individual mass spectrometer calibrations for pure methane alone, and for the apparent methane coming from TMA alone. Note that the m/z = 15 species is greater in magnitude than m/z = 16 resulting from TMA, but the opposite trend is the case for pure CH_4. This is clearly due to the fragmentation of TMA in the ionizer of the mass spectrometer. The difference in slopes, however, means that the species at these m/z ratios may not easily be discriminated as coming from unreacted TMA or from CH_4 reaction product entering the QMS (see text).

measurement, the two QCM crystals, previously coated with Al_2O_3 in the reactor, were measured together for 50 cycles of Al_2O_3. Following this, one of the crystals was removed from the reactor, and a self-assembled monolayer (SAM) was prepared on the crystal by soaking in 1 mM stearic acid (octadecanoic acid) solution in isopropanol for >24 h.[35,36] This highly hydrophobic SAM was expected to inhibit further growth of Al_2O_3 by blocking surface hydroxyl groups, leaving a surface mostly terminated in –CH_3 groups, thus reducing the number of surface sites for TMA chemisorption. The SAM-coated crystal was then re-inserted into the reactor, and an additional 50 cycles of Al_2O_3 were performed. Figure 10 illustrates the difference in growth between these two types of surfaces.

The stearic acid SAM coating clearly inhibits growth significantly for at least 35 cycles, after which the slope of the mass gain with time returns to about the same as the initially uncoated Al_2O_3 crystal. In Figure 10, slight variations

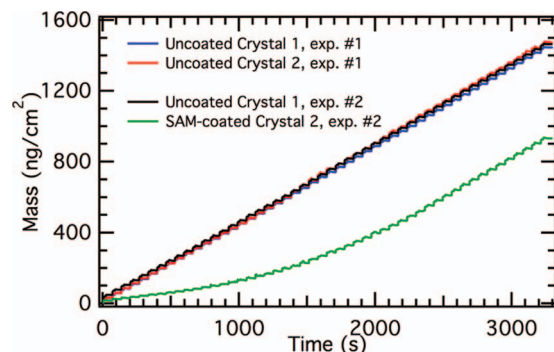on the control sample crystals may be noticed. These are attributed to non-optimal temperature synchronization between the crystals. At high temperatures, fairly precise temperature uniformity over the course of the experiment is necessary to accurately observe small mass differences.[18] One needs to understand the response between the crystal pair when there is no intentional difference to get the best data out of such a comparison and ensure the control experiment results in a more perfect overlap of the top three plots in Figure 10, yet this serves to demonstrate the technique.

Dual QCM experiments have previously been used to study metal depositions by evaporation onto SAMs, and combined with other data to understand the mechanism by which metals adsorb and react with these surfaces.[37,38] In these systems, however, film growth results from metal atom adsorption and condensation onto cool (25 °C) substrates. Hence, absolute sticking probabilities onto SAMs may be calculated by assuming the sticking of evaporants is unity onto the bare metal. With ALD, however, the process is more complicated because of the piecewise cyclic nature of adsorption, and because the adsorption is reactive instead of condensation-like. If the combination of sticking and surface reaction is termed reactive sticking, we can, however, measure a *relative reactive sticking probability* for an ALD precursor pair like TMA/water in this case, with a dual QCM experiment. The sticking probability is relative, because we know that both crystals receive the same integrated flux of precursor (the same exposure), but we are not assuming that the reactive sticking of precursors onto Al_2O_3 is unity. That is to say, whatever the *absolute* sticking probability of precursors is onto the growing film, we are measuring the sticking onto the SAM *relative to* the sticking onto bare Al_2O_3. Figure 11 shows this sticking probability, S(θ), as a function of ALD cycle number, as well as the mass per cycle (MPC) calculated from the data in Figure 10. The S(θ) is calculated by measuring segments of the slopes of the mass gain (from Figure 10)

FIG. 10. Parallel mass measurement onto two QCM crystals. Experiment 1 was 50 Cycles of Al_2O_3 growth from TMA/water at 70 °C onto two crystals previously coated with Al_2O_3 or "uncoated." This was meant to serve as a control. During Experiment 2 the same 50 cycle sequence was performed with a stearic acid self-assembled monolayer (SAM) on crystal 2, and with only bare or uncoated Al_2O_3 on crystal 1. The lower plot shows substrate-inhibited growth, yet the TMA/water reaction is not completely inhibited even on this hydrophobic substrate (see text).
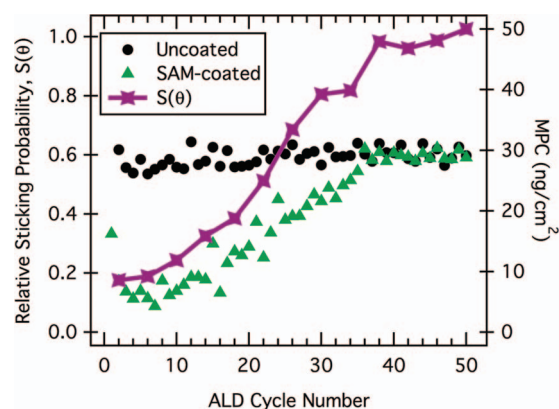

FIG. 11. Relative reactive sticking probability (left axis) and mass per cycle (MPC—right axis) as a function of cycle number. The MPC measurement shows how many cycles are deposited onto the SAM-coated crystal before the same growth rate reappears as on the uncoated crystal. Presumably the SAM is completely covered by Al_2O_3 at that point, which also is indicated by the relative sticking probability returning to ~1. Measurements of this type should lead to a more quantitative understanding of substrate (or surface-chemistry-based) selective growth. TMA is considered an especially reactive precursor and is unlikely to grow completely selectively, yet clearly exhibits some nucleation delay on a methyl-terminated SAM.

for the SAM-coated crystal divided by the slopes of the uncoated crystal, analogously to what was performed for PVD metal deposition onto SAMs.[38]

V. DOSE CONTROL AND GROWTH RATES

With the ability to dose very small quantities of precursor, we can reproducibly grow with either single saturative steps (as is conventionally done) or with single or multiple sub-saturative steps. While it was anticipated that the total MPC ultimately attained should be the same by either method, we have preliminary evidence that the dynamics of the dose, i.e., piecewise vs. all-at-once, matter for the ALD process, in perhaps a way that is subtler than simply whether enough integrated exposure is used to saturate in a single step. It remains under investigation as of this writing whether this is strictly a chemical effect, or a consequence of aspects of the reactor design. By controlling the quantified dose of reactant, relating it to the *net* exposure and *dynamics* of exposure, this reactor design allows a more thorough investigation of the surface chemistry of the ALD process, which will be detailed in a future publication using the unique combination of features and methods herein.

VI. CONCLUSIONS AND FUTURE APPLICATIONS

A unique ALD reactor has been designed and built which allows a new level of precise, accurate, controlled reactant doses. A relationship has been developed between the dose amount and actual impingement amount at the substrate. The combination of *in situ* diagnostics, pumping and flow capabilities, calibrations, and mechanisms to measure the dynamics and totality of each reactant dose's exposure allow the ALD reaction process to be studied in new ways. These include measuring relative reactive sticking probabilities and a proposed complete mass balance on reactant consumption.

Future work with this setup will involve measurement of absolute sticking probabilities of precursors as a function of coverage, and study of the reactant-dose to growth-rate relationships for subsaturative doses. Subsaturative dosing may be useful for highly controlled trace dopant incorporation, potentially leading to improvement in dopant distribution in semiconducting films. It also may potentially be useful for controlling unwanted reactions that oxidize sensitive interfaces, such as using absolutely minimal reagent exposures when depositing gate dielectrics on III–V and other high-mobility substrates, like Ge. The parallel mass measurement approach may be used with barely saturative dosing to explore area selective growth mechanisms, and develop means to design conditions to engineer them.

ACKNOWLEDGMENTS

[1] S. M. George, Chem. Rev. **110**(1), 111 (2010).
[2] R. L. Puurunen, J. Appl. Phys. **97**(12), 121301 (2005).
[3] M. Leskela and M. Ritala, Thin Solid Films **409**(1), 138 (2002).
[4] T. Aaltonen, A. Rahtu, M. Ritala, and M. Leskela, Electrochem. Solid-State Lett. **6**(9), C130 (2003).
[5] T. Suntola, A. J. Pakkala, and S. G. Lindors, U.S. patent 4,389,973 (28 June 1983).
[6] T. Suntola, Thin Solid Films **216**(1), 84 (1992).
[7] B. Sang and M. Konagai, Japn. J. Appl. Phys. **35**(5B), L602 (1996).
[8] J. W. Elam, M. D. Groner, and S. M. George, Rev. Sci. Instrum. **73**(8), 2981 (2002).
[9] D. M. Hausmann, E. Kim, J. Becker, and R. G. Gordon, Chem. Mater. **14**(10), 4350 (2002).
[10] J. S. Becker, E. Kim, and R. G. Gordon, Chem. Mater. **16**(18), 3497 (2004).
[11] R. A. Wind and S. M. George, J. Phys. Chem. A **114**(3), 1281 (2010).
[12] S. T. Christensen, J. W. Elam, F. A. Rabuffetti, Q. Ma, S. J. Weigand, B. Lee, S. Seifert, P. C. Stair, K. R. Poeppelmeier, M. C. Hersam, and M. J. Bedzyk, Small **5**(6), 750 (2009).
[13] J. Lu, B. Fu, M. C. Kung, G. Xiao, J. W. Elam, H. H. Kung, and P. C. Stair, Science **335**(6073), 1205 (2012).
[14] J. W. Elam, Z. A. Sechrist, and S. M. George, Thin Solid Films **414**(1), 43 (2002).
[15] H. Isshiki, Y. Aoyagi, T. Sugano, S. Iwai, and T. Meguro, Appl. Surf. Sci. **82–83**, 57 (1994).
[16] T. S. Christensen, J. W. Elam, B. Lee, Z. Feng, M. J. Bedzyk, and M. C. Hersam, Chem. Mater. **21**(3), 516 (2009).
[17] M. N. Rocklein and S. M. George, Anal. Chem. **75**(19), 4975 (2003).
[18] J. W. Elam and M. J. Pellin, Anal. Chem. **77**(11), 3531 (2005).
[19] D. R. Burgess Jr., J. E. Maslar, W. S. Hurst, E. F. Moore, W. A. Kimes, R. R. Fink, and N. V. Nguyen, AIP Conf. Proc. **788**(1), 141 (2005).
[20] G. Sauerbrey, Z. Phys. A: Hadrons Nucl. **155**(2), 206 (1959).
[21] A. Arnau, Sensors **8**(1), 370 (2008).
[22] J. W. Bender and J. Krim, in *Microscale Diagnostic Techniques*, edited by K. S. Breuer (Springer, Berlin, 2005), p. 227.
[23] L. Donald Smith, *Thin-Film Deposition: Principles and Practice* (McGraw-Hill, New York, 1995), p. 21.
[24] F. W. Sears and G. L. Salinger, *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics*, 3d ed. (Addison-Wesley, Reading, MA, 1975), pp. 254–258.
[25] E. Wiberg, N. Wiberg, and A. F. Holleman, *Inorganic Chemistry*, 1st ed. (Academic, De Gruyter, San Diego, Berlin; New York, 2001), p. 1024.
[26] P. N. Dasgupta, O. Trejo, and F. B. Prinz, J. Vac. Sci. Technol. A **30**(1), 01A110 (2012).
[27] J. E. Maslar, W. S. Hurst, D. R. Burgess, W. Kimes, N. V. Nguyen, E. F. Moore, and J. T. Hodges, ECS Trans. **13**(2), 139 (2008).
[28] B. A. Sperling, W. A. Kimes, J. E. Maslar, and P. M. Chu, J. Vac. Sci. Technol. A **28**(4), 613 (2010).
[29] J. E. Maslar, W. Kimes, and B. Sperling, ECS Trans. **41**(2), 157 (2011).
[30] M. Ritala, M. Juppo, K. Kukli, A. Rahtu, and M. Leskela, J. Phys. IV France **9**(Pr8), 1021 (1999).
[31] M. Juppo, A. Rahtu, M. Ritala, and M. Leskela, Langmuir **16**(8), 4034 (2000).
[32] I. J. Hsu, D. A. Hansgen, B. E. McCandless, B. G. Willis, and J. G. Chen, J. Phys. Chem. C **115**(9), 3709 (2011).
[33] L. A. Belfiore, *Transport Phenomena for Chemical Reactor Design* (Wiley, New York, 2003).
[34] Jurgen H. Gross, *Mass Spectrometry: A Textbook*, 2nd ed. (Springer, Berlin, 2010).
[35] D. L. Allara and R. G. Nuzzo, Langmuir **1**(1), 45 (1985).
[36] D. L. Allara and R. G. Nuzzo, Langmuir **1**(1), 52 (1985).
[37] T. B. Tighe, T. A. Daniel, Z. Zhu, S. Uppili, N. Winograd, and D. L. Allara, J. Phys. Chem. B **109**(44), 21006 (2005).
[38] A. V. Walker, T. B. Tighe, O. Cabarcos, B. C. Haynie, D. L. Allara, and N. Winograd, J. Phys. Chem. C **111**(2), 765 (2007).