Metrology, Modeling, and Control

EECS 290H   Special Issues in Semiconductor Manufacturing

Profs. Costas Spanos & Kameshwar Poolla

Fall 2003

one model-based control

- why use it ? how does it help ?
- ingredients
- control architectures

two sensors

- general issues and tradeoffs
- temperature sensors (pyrometry, thermocouples, acoustic wave)
- pressure and flow sensors (manometers, McLeod transducer, momentum devices)
- composition sensors (OES, LIF, RGA, Mass Spectroscopy, actinometry)
- thickness sensors (reflectometry, ellipsometry)
- post-process sensing (SEM, AFM, other microscopy)

three   filtering, estimation, modeling

- introduction
- control-oriented modeling
- kalman filtering
- an example : reflectometry
- parameter estimation

four   run-to-run control

- introduction
- a simple scheme
- analysis
- case study : resist thickness control

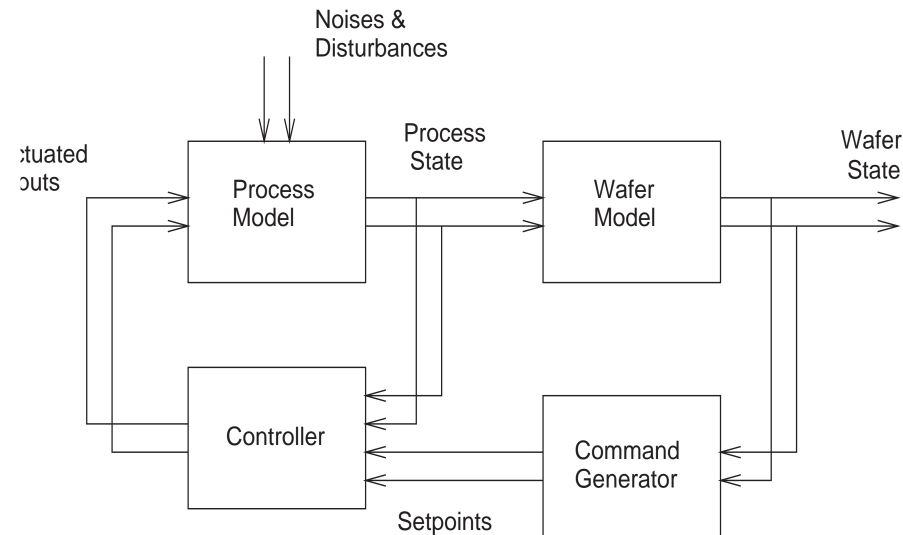
five   real-time control

- introduction
- case study : rapid thermal processing
- case study : reactive ion etch
- implementation issues

- Why do control ?
  - control is an *enabling technology* for the next generation of integrated electronics
  - control is a cost-effective means of *retrofitting* existing fab lines
  - importance is gaining acceptance – Industry, UCB, Stanford, Michigan, CMU, Texas

- How does control help ?
  - can reduce variability in *product parameters* (ex: CD, sidewall, etc.) by regulating *process parameters* (ex: etch rate, bake temp, etc.)
  - can reduce time to re-calibrate a process (ex: etchers taken off line for routine cleaning)
  - required for effective sensor development
  - can provide early diagnostic warnings (ex: control signals)
  - can accelerate time-to-yield
  - result: higher yields & equip utilization, lower product variance with modest capital cost

- Ingredients of control
  - *model* : control-oriented, need not be detailed or accurate
  - *sensors* : to measure process parameters
  - *actuators* : to change process operating point
  - *reference command* : desired value (or trajectory) of process parameters
- Hierarchical control levels
  - *supervisory* : oversees commands for wafers lot-by-lot, uses SPC and response surface models,
  - *run-to-run* : command for a wafer are generated based on metrology of previous wafer
  - *real-time* : requires modern control and ID methods, accurate in situ metrology



- controller keep process parameters constant even with disturbances and using a coarse model
- this results in product parameters being regulated to constant values
- we can later map the process parameters to the product parameters to generate appropriate reference commands

two sensors

- general issues
- tradeoffs
- temperature sensors (pyrometry, thermocouples, acoustic wave)
- pressure and flow sensors (manometers, McLeod transducer, momentum devices)
- composition sensors (OES, LIF, RGA, Mass Spectroscopy, actinometry)
- thickness sensors (reflectometry, ellipsometry)
- post-process sensing (SEM, AFM, other microscopy)

- Sensors (and actuators) are key limiting factors in application of control techniques to semiconductor manufacturing
- sources of difficulty
  - implementation environment (vacuum, clean facilities, etc.)
  - perception that in-situ sensors affect process
  - ex-situ sensors can reduce throughput
  - cost of ownership
  - traditional resistance in industry

- modeling is often key part of sensing
  - physical quantity of interest may not be directly measured (ex: OES indirectly contains info about etch process state)
  - thus, sensors are based on a model of the underlying physical process

$$sensors = model + data$$

- signal processing
  - needed to reduce noise, improve bandwidth
  - difference between *data* and *information*

- problems
  - sensors require calibration
  - must account for drift

- other issues
  - sensor fusion
  - data compression

- some key tradeoffs
  - non-invasive vs. invasive
  - non-destructive vs. destructive
  - in-situ vs. ex-situ
  - speed vs. accuracy
  - noise
- bias (accuracy) vs. variance (repeatability)
  - a sensor could be inaccurate,
    (ex: a thermocouple readings are off by $4^\circ K$)
  - but the sensor might have good repeatability, (ex: it is consistently off)
  - repeatability is often more important for process control
- modern filtering and estimation methods can be of great use in improved sensing software. (an example from reflectometry later)

- operating principle

  Peltier-Seebeck effect, up to $3000°C$

  T gradient along wires of different materials develop different emf

  emf measures junction T

  platinum rhodium alloy, or silicon based

  sensitivity $100 - 200\mu V/°K$

- problems

  big problems with shield design

  radiative effects

  low signal – need amplifiers or use thermopile

  invasive

  gas T measurement is very hard, especially $< 10^{-4}$ torr

- comments

  inexpensive, low drift

  accuracy $\approx \pm 5°C$ at $800°C$

  low bandwidth

  where do you want to measure T ?

- operating principle

  acoustic wave is transmitted through body

  surface and internal waves propagate thru body at T dependent speed

  interference with source gives beats

  beat frequency determines T

- issues

  implementation difficulty

  invasive

  calibration

- operating principle

  hot objects radiate

  radiation is wavelength dependent

  radiation model for black bodies (Planck's Law)

  $$R_\lambda = \frac{37418}{\lambda^5 \left(e^{14388/\lambda T} - 1\right)}$$

  λ in $microns$, T in °K, R_λ

  for non-black bodies need to account for $emissivity$

- issues

  surface properties affect radiation

  multiple internal reflections

  emissivity is wavelength and geometry dependent

  can change during processing

  calibrations via thermocouples, difficult

- direct gauges

  displacement of a solid or liquid surface

  capacitance manometer, McLeod pressure transducer

- indirect gauges

  measurement of a gas related property

  momentum transfer, charge generation

- huge range of available sensors

  cost

  sensitivity

  range

- basic idea

  pressure differential causes displacement of diaphragm

  sense capacitance change between diaphragm and fixed electrode

  resolution 10^{-2} percent at 2 hertz and 10^{-3} torr

- differential pressure meters
- thermal mass flow meters

  mass flow $= \dfrac{K}{T_1 - T_2}$

  K depends on specific heat of gas etc.

  must be calibrated for different gases

  accuracy ≈ 1 sccm at flows of 40 sccm

  low bandwidth because of thermal inertia

- measures concentration of various species present in plasmas
- useful in various plasma etch and plasma-enhanced deposition control applications
  - endpoint detection
  - impurity detection
  - etch rate monitoring
  - uniformity measurement
- provides real-time measurements (> 1 Hz)
- simple installation on most plasma etchers.

- operation principle
  - plasmas contain ions, neutral radicals, energetic electrons
  - plasma discharge light

$$A + e \longrightarrow A^* + e, \quad A^* \longrightarrow A + h\nu$$

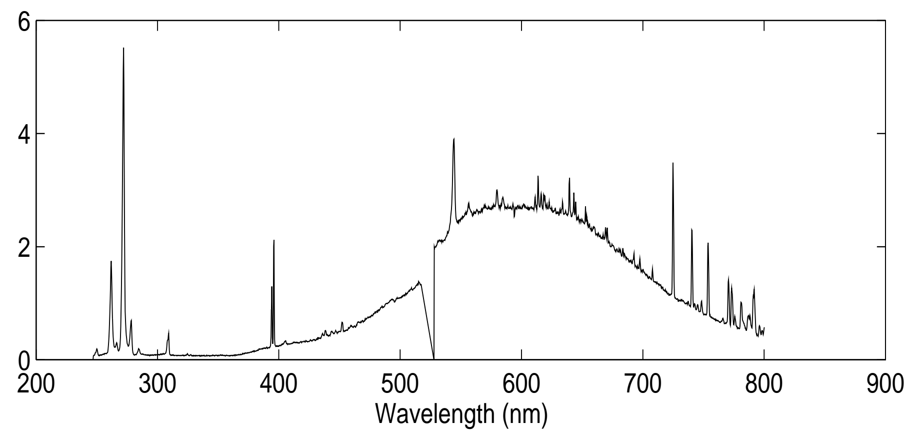
here A^* is the excited state of particle A

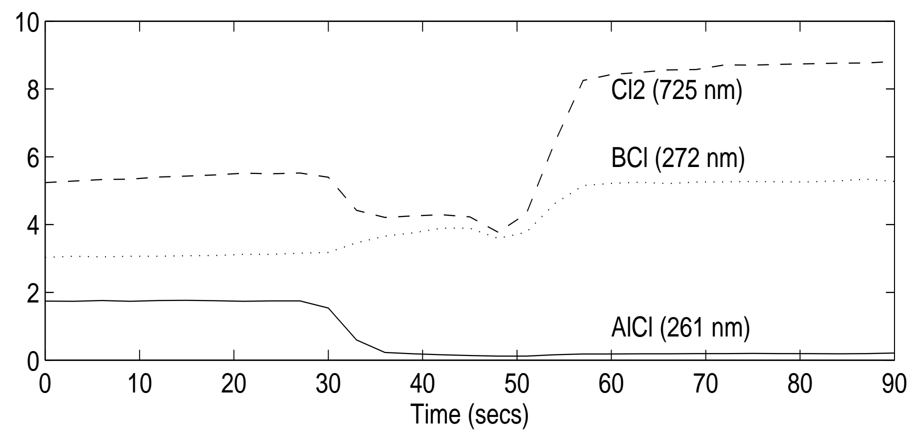
  - frequency of emitted light

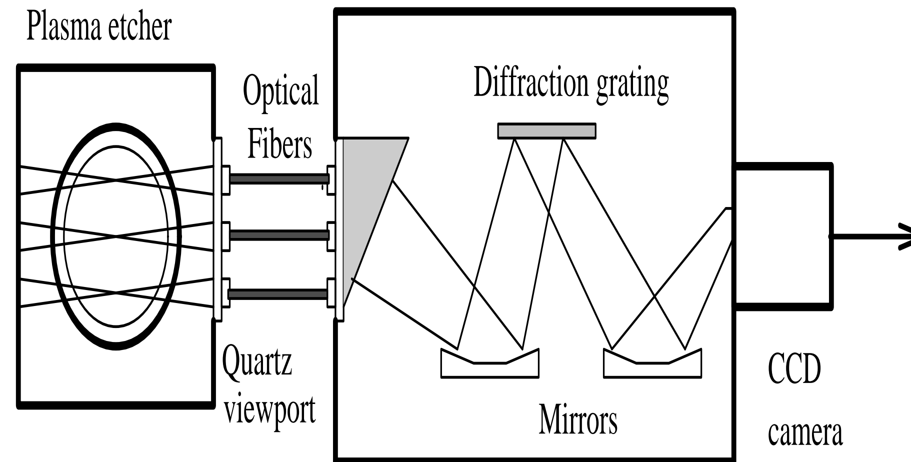
    depends on allowable energy transitions

    is characteristic of species

    sometimes there is no useful emission signature in OES
    (ex: SiH_3 in PECVD with silane plasma)


6
4
2
0
200
300
400
500
600
700
800
900
Wavelength (nm)


10
8
6
4
2
0
0
10
20
30
40
50
60
70
80
90
Time (secs)
Cl2 (725 nm)
BCl (272 nm)
AlCl (261 nm)



- optical equipment options
  - photo-detector (possibly with scanning of diffraction grating)
  - photo-diode array
  - CCD camera
  - Choice depends on number of factors
    frequency resolution, spatial resolution, acquisition rates,
    bandwidth, sensitivity, etc.

- signal Processing Issues
  - OES intensities depends on several factors in addition to species concentration, such as

    Excitation probability (strongly dependent on RF power),

    Optical collection efficiency (drifts over time due to residue build-up on window).
  - full-spectrum OES may require data compression and noise reduction.
  - signal intensity may be too weak in

    Small area etches (vias and contact cuts in oxide)

    Detection of trace Cu sputter targets in Al-Cu etches.

- basic idea

  use a pulsed laser to excite plasma, observe induced emission

  laser can be tuned to cause specific excitations

  can detect species that have no natural emission (ex: SiH_3)

  can detect species in ground state

- details

  $Nd-YAG$ laser source @1064 nm used to pump tunable dye laser

  pulsed lasers provide much more power in short excitation phase
  thus emission exceeds background

  collection optics at 90° to source to minimize scattered light

  can detect and measure $CF, CF_2, SiO, SiN, BCl, Cl_2+, ...$

- issues

  excellent spatial resolution (5 microns)

  excellent temporal resolution (100 nsec)

  sensitivity $10^6 - 10^8$ particles/cm^3

  complex collection optics and signal processing

  more complex than OES, but much more accurate

  requires side viewing port

  requires actinometry for calibration

  limited to species with absorption in $200 - 900$ nm range

- other option – Laser absorption spectroscopy

  tunable laser diode is used as source in IF range

  absorption is very low here, so multiple passes are needed

  path length is ≈ 1 Km

  poorer resolution

  qualitative tool

- objective: calibrate OES/LIF signals
- basic idea

  introduce known amount of inert gas B (ex: Ar)

  choose wavelength in inert gas emission spectrum whose excitation X-section, and excitation energy resembles species of interest A. Then,

  $$\frac{I_a(\lambda_a)}{I_b(\lambda_b)} = K\frac{N_a}{N_b}$$

  I_a, I_b measured intensities
  $N_i =$ molar fraction of input gas i
  $K =$ actinometric constant (assumed fixed for other species also)

- issues

  useful only for measurement *relative* species concentration

  repeatability is a big issue: must ensure that emission lines go thru same optical path, uniform temporal electron densities, etc.

- used to spectrally resolve light
- operating principle

  close parallel lines or steps etched on a surface

  mechanically made gratings : etched glass or plastic

  holographically patterned gratings : higher transmission, flatter response

  modern gratings are blazed : periodic phase shifting across grating, concentrates light energy in a specific order

- performance characteristics

  peak location is at $sin\theta = m\lambda/d$

  resolving power $R = \dfrac{\lambda}{\Delta\lambda} = Nm$

  dispersion $D = \dfrac{\Delta\theta}{\Delta\lambda} = \dfrac{m}{d\cos\theta}$

  Here, d = spacing, N = number of lines, m = order
  λ = wavelength of incident light, θ = viewing angle

- basic idea
  - essentially a tunable narrow-band wavelength selective optical filter
  - uses a diffraction grating
- issues
  - accuracy of selected wavelength
  - calibration
  - efficiency (transmission $\approx$ 10 %)
- Czerny-Turner monochromator
  - grating is rotated by a stepper drive
  - angle of rotation determines wavelength of light at exit slit
- dielectric bandpass filters
  - fixed wavelength applications
  - transmission $\approx$ 50 - 70 %

- photo-multipliers

  very high gain photon detectors – rely on cascading

  resolution 0.05 photons/sec

  spectral characteristic are adjustable by choice of material

- photo-diode arrays

  can be used directly to measure intensity vs. wavelength

  lower resolution than a monochromator with pmt

  wider spectral coverage than monochromator with pmt

  light strikes a multi-channel intensifier plate and emits electrons

  DC bias accelerates electrons towards a phosphor target

  fiber optics connect to pixels (up to 1200)

  need to be cooled to limit thermal photo-electron emissions

  smaller dynamic range

- CCD arrays

  similar operating principle as photo-diode arrays

  no multi-channel intensifier plate

  wider spatial coverage

  lower resolution

  very inexpensive

- operating principle

  interference between light reflected from top surface and
  from surfaces in underlying stack

  intensity at detector is $\approx$ sinusoidal in thickness of top layer and in wavelength

  provides thickness/index/composition measurement of top layer

  typically near-normal incidence

- applications

  etch-rate measurement

  develop-rate measurement

  end-point detection

- single-wavelength

  pulsed laser source is preferred

  detected light is filtered at pulsing frequency to reduce noise

  thickness accuracy $\pm 20\mathring{A}$

- spectral reflectometry

  broad band incoherent source for scan-wavelength

  more noisy, but can solve for more unknowns (index and composition)

- issues

  absolute intensity measurement is hard, need to model optics

  need underlying stack geometry and indices

  multiple internal reflections

  phase shifts at stack boundaries

  surface roughness

  patterned wafers

- operating principle

circularly polarized incident light

$\perp$ and $\|$ components undergo different reflections

Fresnel reflection coefficients

$$r_{12}^{\perp} = \frac{n_2 \cos\phi_1 - n_1 \cos\phi_2}{n_2 \cos\phi_1 + n_1 \cos\phi_2}$$

$$r_{12}^{\|} = \frac{n_1 \cos\phi_1 - n_2 \cos\phi_2}{n_1 \cos\phi_1 + n_2 \cos\phi_2}$$

received light is elliptically polarized

- issues

    polarization of detected light is measured by nulling

    no need to have intensity measurements

    polarized light source

- comments

    can measure 2 quantities – typically index and film thickness

    spectrally and spatially resolved ellipsometry

    extremely accurate – index $\pm.1\%$, thickness $\pm4\mathring{A}$

    more expensive and delicate than reflectometry

- two types

  flux analyzers : sample gas thru aperture

  partial pressure sensors : analysis in exhaust stack

- issues

  recombination in mass spec tube changes

  indistinguishable species : (ex: CO, N_2 and Si have same amu (28))

  pressure measurements are removed from processing chamber

- basic idea

  special kind of mass spectrometer

  measures gas compositions

  works at low vacuum $< 10^{-5}$ torr

  ion beam is produced from gas sample by e-bombardment

  beam is collimated by electric fields

  q/m ratio of ions determines bending in B field

  detection of ions via a Faraday cup

- issues

  quadrupole (magnetless design)

  very noisy !!

  good for diagnostics

  can withstand $500°C$

  can also be used at higher pressures with differential pumps

  mass range $50amu$, resolution $2amu$,

- microscopy
  - optical
  - scanning electron microscope
  - atomic force microscope
- other tools
  - resistivity measurements for CD
  - stress measurements

three modeling, estimation, filtering

- introduction
- control-oriented modeling
- parameter estimation
- kalman filtering
- an example : reflectometry

- modeling objectives
  - simulation studies
  - equipment/process design
  - process control
  - sensor development
  - diagnostics & fault detection
- different objectives call for different kinds of models
  - phenomenological models

    very accurate, detailed

    mass and energy transport, FEM based

    do not run in real-time

    needed for equipment design, process simulation
  - control-oriented models

- properties
  - simplified, low order
  - gives first-order trend information
  - may run in real-time
  - used for control and diagnostics
- methods
  - black-box
  - grey-box
  - model reduction
- issues
  - experiment design
  - choice of model structure
  - choice of ID procedure
  - model bias and variance
  - model verification

- ideas
  - model will have good prediction on inputs with spectral content similar to those used in ID
  - averaging to reduce effects of noise

$$\text{variance in estimated parameters} \approx \frac{\text{noise variance}}{\text{reduction factor}} = \frac{\sigma_e^2 * n_p}{L * n_e}$$

L	number of data points
n_e	number of noise channels
n_p	number of parameters
σ_e^2	noise variance

  - model verification

    use fresh input-output data set

    check residuals : white ? uncorrelated with input ?

- parametric

  transfer functions

  ARX, ARMAX

  general parametric

  required for any computation

- non-parametric

  frequency response

  impulse / step response

  wiener kernels

  need model reduction to get parametric model before control is done

- dealing with nonlinearities

  gain-scheduled models

  semi-local nonlinear models (Hammerstein, Weiner, etc.)

- basic idea

  take DFT of input u and output y

  $$H(j\omega) = \frac{Y(j\omega)}{U(j\omega)}$$

  computed at N frequencies where $N =$ length of data record

  bias $= 0$ (asymptotically)

  variance $= \dfrac{\sigma^2_{e(\omega)}}{\sigma^2_{u(\omega)}}$

  estimates at different frequencies are uncorrelated

- issues

  smoothing the estimate, choice of smoothing window

  bias vs. variance

  converting to a parametric model

  nonlinearities, time-variations (under-modeling)

- basic idea

  state-space model

$$
\begin{aligned}
x_{k+1} &= Ax_k + Bu_k \\
y_k &= Cx_k + Du_k
\end{aligned}
$$

  model parameters A, B, C, D obtained by factoring a certain matrix computed from data

  works very well for multi-variable systems

  no proven optimality properties

  can get order estimates also

- issues

  loss of physically meaningful parameters

  computationally intense: cannot be done in real-time

  nonlinearities, time-variations (under-modeling)

- general modeling problem
  - given: data u^d, y^d, model $y = f(u, \theta, e)$
    where e is unit-variance Gaussian white noise
  - find: "best" estimate of parameters θ
- solution strategy: (maximum likelihood estimation)
  - note that for any fixed θ, y is a random vector.
  - compute the density function $p_Y(y; \theta)$. Note that

    $$p_Y(y^d; \theta) \;\propto\; \text{the likelihood that } y = y^d, \text{ given parameter values } \theta$$

  - good choice of parameters

    $$\begin{aligned}\theta_{ML} &= \text{arg} \max_\theta \; p_Y(y^d : \theta) \\ &= \text{arg} \min_\theta J(\theta)\end{aligned}$$

    where $J(\theta) = -\log p_Y(y^d : \theta) =$ log-likelihood function.

- computing the estimate: nonlinear programming
- can easily

  incorporate prior information (statistics on θ)

  compute variance of parameter estimates
- issues

  can only realistically treat gaussian noise case

  nonlinear models are linearized, and then

  resulting parameter estimates can be biased

  choice of model structure f is key : basic physics/chemistry

  too many parameters is bad

  sensitive to noise model

- problem:

  Let y_k be independent random variables jointly Gaussian as $\sim \mathcal{N}\left(\theta, \sigma_k^2\right)$
  known variances σ_k, unknown (common) mean θ
  Given observations y_k^d for $k = 1, \cdots, L$
  estimate the mean θ

- solution:

  Log-likelihood function is

$$J(\theta) = \sum_k \frac{\left(y_k^d - \theta\right)^2}{2\sigma_k^2}$$

  This is a quadratic and is easy to minimize. We get

$$\theta_{ML} = \sum_k y_k^d w_k, \quad \text{where} \quad w_k = \frac{\sigma_k^2}{\sum_\ell \sigma_\ell^2}$$

  natural interpretation.

- consider a linear system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + B_u u(t) + B_v v(t) \\ y(t) &= Cx(t) + D_u u(t) + w(t) \end{aligned}$$

$x =$ process state

$v =$ process noise : used to account for under-modeling

$w =$ measurement noise

$u =$ process input

- assumptions

$w(t), v(t), x(0)$ uncorrelated, known mean and covariance

- problem

build a box (in software) that processes $y(t)$ and $u(t)$ and produces an estimate $\hat{x}(t)$ of the state

- mathematically,

$$\begin{aligned} \hat{x} &= F(u, y) \\ F_{opt} &= \arg \min_{F} \mathbf{E} (\hat{x} - x)(\hat{x} - x)^* \end{aligned}$$

- optimal filter realization
  - simple structure

$$\dot{x}(t) = A\hat{x}(t) + B_u u(t) + K(t)\nu(t)$$

  - copy of model driven by input u and innovations

$$\nu(t) = (y(t) - C\hat{x}(t) - D_u u(t))$$

  - kalman gain $K(t)$ computed by solving a Riccati differential equation
  - for linear time-invariant problems $K(t) \to K_\infty$
    found by solving a algebraic riccati equation
- optimality in what sense ?
  - for gaussian noises, optimal over *all* filters
  - for general noises, optimal over *all linear* filters

- kalman gain $K(t)$
  - captures tradeoff between model accuracy (process noise) and sensor noise
  - results in white innovations, uncorrelated with output $y(t)$
- easy recursive implementation
  - requires one linear system simulation
  - requires solving one riccati equation
- very general problem
  - recursive least squares is a special case
  - can also do parameter estimation in same framework
- nonlinear systems ?
  - *Extended Kalman Filtering*
  - basic idea: kalman filter for the linearized system
  - not optimal in general
  - optimal nonlinear filtering: eextremely difficult

- basic idea
  - significant improvement in reflectometry sensor for etch-rate estimation by incorporating a model and using EKF
  - conventional etch-rate metrology
- setup
  - fix a wavelength λ (analysis can be done for multiple λ)
  - reflected light intensity I_r is

$$\begin{aligned} I_r &= I_o r(d, \lambda, \phi_1, \cdots, \phi_n) \\ r &: \text{reflection coefficient} \\ d &: \text{film thickness} \\ \phi_1, \cdots, \phi_n &: \text{index and thickness parameters for underlying stack} \end{aligned}$$

  - measurements

$$y = \alpha I_r + v = \alpha f(d) + v$$

where v is meas. noise, and α captures effect of optics

- conventional methods for etch rate estimation
  - peak counting

    insensitive to noise, easy implementation

    gives average, not instantaneous, etch rate

    difficult to merge information from multiple wavelengths
  - nonlinear least squares

    solve at each time

    $$\min_{d} \ \|y_k - f(d_k)\|^2$$

    computationally expensive

    etch rate is estimated by numerical differentiation
- EKF based method

  gives direct etch rate estimate

  computationally reasonable

  can merge multiple wavelength data easily

- state-space model used

$$\begin{aligned} d_{k+1} &= d_k - T er_k \\ er_{k+1} &= er_k + w_k^{(1)} \\ \alpha_{k+1} &= \alpha_k + w_k^{(2)} \\ y_k &= \alpha_k f(d_k) + v_k \end{aligned}$$

here, er is etch rate, T is sampling time

drift of etch rate and optical gain is modeled as a random walk

- EKF details

  state estimate gives er estimate

  Ekf requires linearizing model, i.e. gradient of f

  also need to solve riccati eqn

  riccati eqn has two knobs Q, R : measurement and process noise variance

  these were tweaked to get good agreement with
  independent etch rate measurements

four   run-to-run control

- introduction
- issues
- an example
- analysis
- case study : resist thickness control

- motivation:
  compensate for process drift
- basic idea:
  - build model from process inputs to measured outputs
  - measure process/wafer parameters for k^{th} wafer
  - use measurements to adjust process input settings for $k+1^{\text{st}}$ wafer
    this requires the model built earlier
  - update model parameters as needed
- result:
  wafer variability is reduced while increasing the variability at process input
- performance measures:
  - within lot wafer parameter bias & variance
  - lot-to-lot wafer parameter bias & variance

- objections:
  - E. Deming "Out of the crisis : quality, productivity and competition", Cambridge, 1986
  - RTR (or any other feedback) increases process variance
  - RTR control can hide process / equipment defects, failures, drifts.
    This will eventually lead to catastrophic failures
  - RTR control requires expensive modeling, design, hardware
- responses:
  - This is true if sensor noise exceeds process variability.
    In many situations, RTR control can reduce variance.
    (more precise statement later)
  - RTR control does not hide process changes.
    The controlled input settings will reveal process drifts.
    No information loss
  - Expense is justified only after a cost/benefit analysis.
    For new technologies, control is *necessary*.

- issues:
  - choice of model structures
    - neural networks
    - parametric models
    - time-series models
    - adaptation of models
  - choice of control strategies
    - fuzzy logic
    - PID control
    - gain-scheduled control
  - choice of modeling method
    - neural training
    - classical non-parametric ID
    - parameter estimation

- open problems:
  - relationship to SPC

    SPC and RTR use same measurements

    in SPC, key outputs are monitored and we detect
    if a process is outside control limits
  - quantifying & measuring performance
- what would be really nice :
  - get a few numbers from the process

    sensor bias & variance

    process bias & variance

    process drift
  - use this info to

    decide if RTR is worthwhile

    design RTR (forgetting factors, control gain)

- problem setup
  - suppose the process output can be modeled as $y_k = \alpha_k + \beta_k u_k + e_k$

k	wafer index
u_k	input setting
α_k, β_k	model parameters
e_k	cumulative effect of noise/ disturbances

  - note that we allow α, β to depend on k to incorporate process drift
- objective
  - we wish to regulate y to the constant reference value r
- certainty equivalent adaptive control
  - if we knew $\alpha_{k+1}, \beta_{k+1}$, we would use the input setting $u_{k+1} = \dfrac{r - \alpha_k}{\beta_k}$ for the next wafer
  - since we don't know the model parameters, we use our best estimates $\hat{\alpha}_{k+1}, \hat{\beta}_{k+1}$ based on past data

- estimating the parameters
  - now suppose β is constant and known
  - one estimate for α is

    $$\hat{\alpha}_{k+1} = (1 - w)\hat{\alpha}_k + w(y_k - \beta u_k)$$

    here, w is a weight that discounts old data
- more generally, we can use a
  kalman filter to recursively estimate the parameters
  - automatically tradeoff parameter drift and measurement noise
  - need Q, R, P_0 : drift variance, sensor variance, initial guess variance
- analysis
  - performance analysis of certainty equivalence adaptive control
    is hard in general
  - can study simple cases analytically
  - more complex cases by simulation

- stability analysis of simple example
  - suppose true plant model is $y_k = \alpha_k + bu_k + e_k$
  - closed loop behaviour is

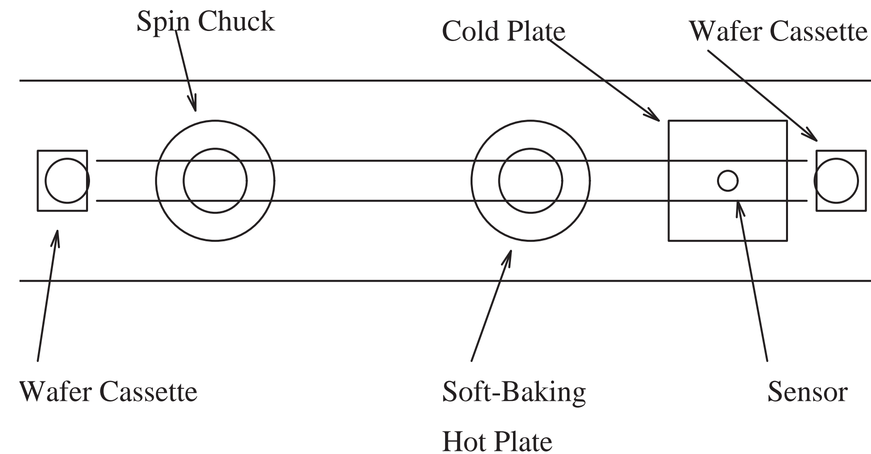
$$\begin{aligned} u_{k+1} &= u_k - \frac{w}{\beta}(y_k - r) \\ &= \left(1 - \frac{wb}{\beta}\right) u_k - \frac{w}{\beta}\left(\alpha_k + e_k - r\right) \end{aligned}$$

  - *integral* action control
  - closed-loop behavior is stable $\Longleftrightarrow$ $0 \leq \dfrac{wb}{\beta} \leq 2$
  - thus, we need sign of b correctly

- performance analysis of simple example
  - assume stability
  - since closed-loop is a linear system, can analyze r and e separately
- convergence analysis
  - first assume $\alpha = \alpha^{\circ} =$ constant, $e = 0$ (no noise)
  - then $(y_k - r)$ decays to zeros exponentially as $\left(1 - \dfrac{wb}{\beta}\right)^k$
  - for rapid response, need $wb = \beta$, or $w \approx 1$
  - robustness analysis
- noise analysis
  - it happens that $\sigma_y^2 = \dfrac{2\sigma_e^2}{2 - w}$
  - Deming is proved right !?
  - true if e is the only source of variability
  - if α also varies, RTR can reduce σ_y^2

- lithography product parameters
  - critical dimension & overlay error
- affected by
  - resist thickness & uniformity
  - PAC concentration
  - exposure dose, stepper vibration and alignment
  - pre-bake, post-bake, & develop times
- we focus on controlling the resist coating and post-bake process. Objective is to remove process variation in
  - resist thickness & PAC concentration
- motivation
  - resist thickness determines focal plane setting
  - PAC determines exposure dose

- equipment used
  - SVG 8626 wafer track
  - 8363 HPO bake plate
  - OCG-820 resist
- processing steps
  - pre-bake and HMDS
  - resist deposition
  - post-bake, cool, measure



- measured variables are
  - resist thickness t
  - photo-active compound concentration PAC
- metrology details
  - indirect measurement of t, PAC
  - use Xenon light source and photospectrometer to get reflectance vs. wavelength curves
  - software to curve fit in two separate bands to get t and PAC.
  - need thickness and index (or composition) of underlying layer

spin coating process is most significantly affected by

- spin speed ω
  - increased ω results in decreased t
  - increments limited to ± 100 rpm
- bake Temperature T
  - solvent evaporation reduces t
  - baking reduces sensitivity of PAC
  - increments limited to $\pm 3^\circ$ C

This is why we choose to actuate ω and T

- in addition to actuated inputs, process is affected by
  - resist viscosity
  - environmental effects
  - wear and age of wafer track
- we regard these as unmodeled disturbances.
- could have used an environmentally controlled chamber – expensive.

- to regulate resist thickness t, we need a model
  - simple log-linear model in literature

$$\log t = a \log(\omega) + b + w$$

  - w is measurement noise, a, b are constants
- could have used a complex model
  - good for simulation
  - bad for control as controller complexity $\approx$ model complexity
- simple model has poor predictive capability across a lot
  - does not include process disturbances
  - to do this, we let model parameters drift slowly

$$\begin{aligned} \log t_k &= a_k \log(\omega_k) + b_k + w_k \\ \begin{bmatrix} a_{k+1} \\ b_{k+1} \end{bmatrix} &= \begin{bmatrix} a_k \\ b_k \end{bmatrix} + v_k \end{aligned}$$

  - v_k is process noise to account for disturbance effects

- standard assumptions
  - v_k, w_k are zero-mean, Gaussian, white, independent
  - (unknown) covariance matrices $Q = E[v_k v_k']$ and $R = E[w_k w_k']$
- to simultaneously regulate t and PAC we use a similar model

$$
\begin{aligned}
\log t_k &= a_k \log(\omega_k) + b_k \log(T_k) + c_k + w_k^{(1)} \\
\log PAC_k &= d_k \log(\omega_k) + e_k \log(T_k) + f_k + w_k^{(2)}
\end{aligned}
$$

$$
\begin{bmatrix} a_{k+1} \\ \vdots \\ f_{k+1} \end{bmatrix} = \begin{bmatrix} a_k \\ \vdots \\ f_k \end{bmatrix} + \mathbf{v}_k
$$

- model above
  - is empirical
  - experiments were conducted to validate assumptions
  - model is adequate for small input changes

- standard Model form

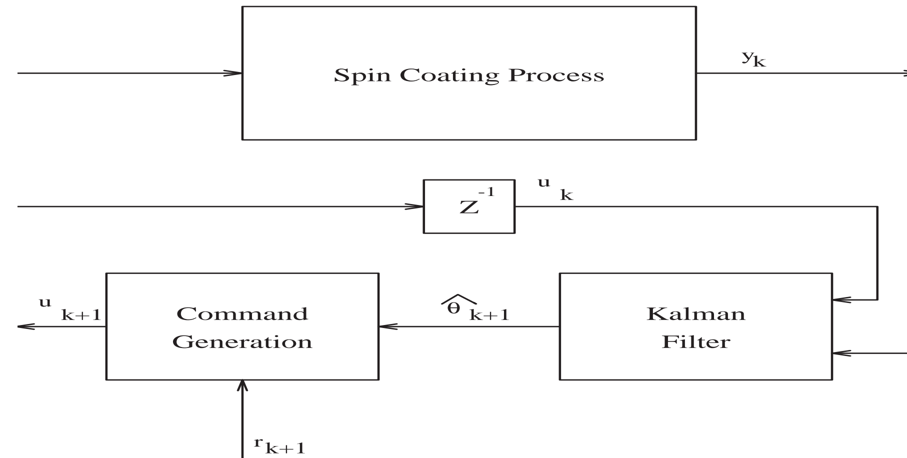
$$\begin{aligned} y_k &= C_k\theta_k + w_k \\ \theta_{k+1} &= \theta_k + v_k \end{aligned}$$

- kalman filter permits optimal recursive estimation of parameters θ_k based on past measurements

$$\begin{aligned} \hat{\theta}_{k+1} &= \hat{\theta}_k + L_k\left(y_k - C_k\hat{\theta}_k\right) \\ L_{k+1} &= P_{k+1}C'_{k+1}\left(C_{k+1}P_{k+1}C'_{k+1} + R\right)^{-1} \\ P_{k+1} &= P_k - L_kC_kP_k + Q \end{aligned}$$

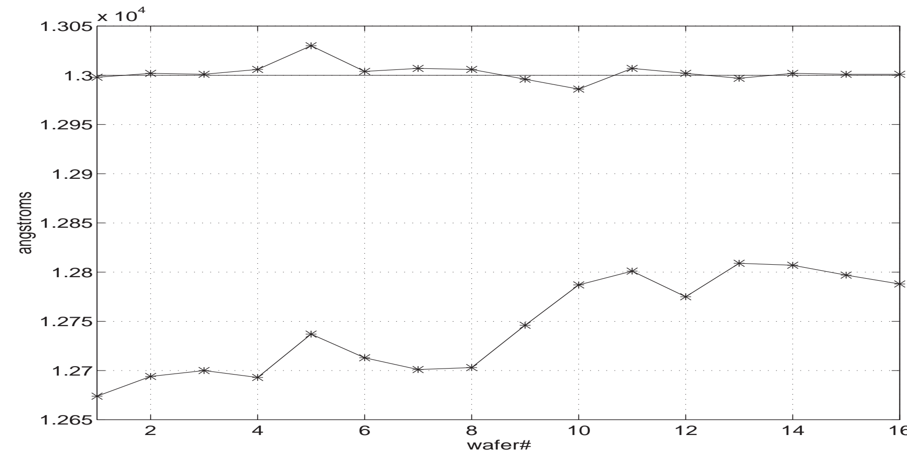
- need three inputs to initialize filter
  - $Q = E\left[w_kw'_k\right]$
  - $R = E\left[v_kv'_k\right]$
  - $P_0 = E\left[\theta_0\theta'_0\right]$

Kalman-Filter based Adaptive Command Generation



- basic Idea:
  - use past measurements to estimate current parameter $\hat{\theta}_k$
  - estimator discounts old data
  - generate new command ω_{k+1} based on model and desired thickness r
  - User specified inputs: (machine dependent)
    sensor variance Q
    within-lot parameter variance R
    initial parameter variance P_0

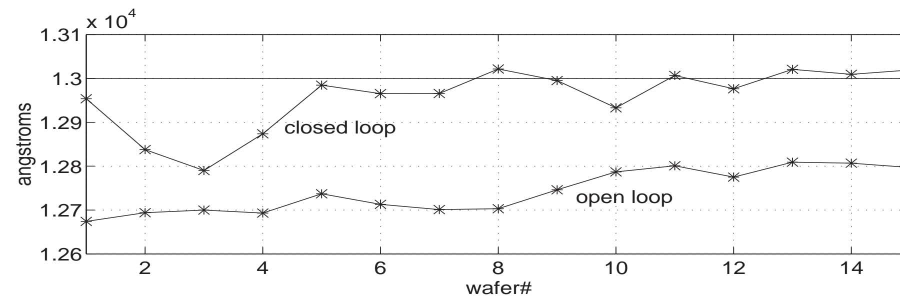
- experimental details
  - bake temperature held constant to suppress its effects
  - unpatterned wafers SiO_2 on Si substrate
  - SiO_2 layer was $1188 \pm 6\mathring{A}$
  - prebake and HMDS omitted
- sensor variance R obtained by repeated measurements on a coated wafer at same location $R \approx 27\mathring{A}$
- process variance Q obtained by minimizing prediction error over a wafer run of 15 wafers
- initial parameter values $\hat{\theta}_0$ and variance P_0 obtained by optimal fitting for each run and sample-path averaging over a 15 wafer run

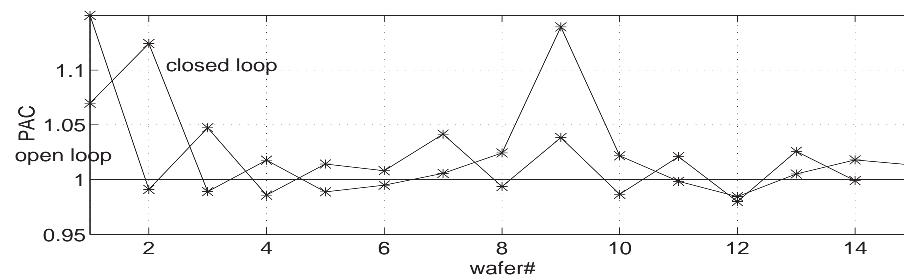


- Open Loop
  - offset from target (lot-to-lot variance) $\approx \pm 300\mathring{A}$
  - wafer-to-wafer variance $\approx \pm 60\mathring{A}$
- Closed Loop
  - offset from target (after two wafers) $\approx \pm 30\mathring{A}$
  - wafer-to-wafer variance $\approx \pm 30\mathring{A}$
    theoretical limit = sensor variance $(27\mathring{A})$

- sensor variance R
  - for t obtained as before $\approx 27\mathring{A}$
  - for PAC we took repeated measurements on a wafer processed with $\omega = 4600$rpm and $T = 90^\circ$ C ≈ 0.012 (normalized units)
- process variance Q, and initial parameter variance P_0 obtained as before

	Open Loop	Closed Loop
bias in t	$300\mathring{A}$	$30\mathring{A}$
variance in t	$60\mathring{A}$	$45\mathring{A}$
bias in PAC	$.04$	$.04$
variance in PAC	$.03$	$.04$


x 10^4
1.31
1.3
1.29
1.28
1.27
1.26
angstroms
closed loop
open loop
2
4
6
8
10
12
14
wafer#


1.1
1.05
1
0.95
PAC
closed loop
open loop
2
4
6
8
10
12
14
wafer#

- outlier: due to actuator limit on changes in T to $\pm 3^\circ$ C
- need to run more wafers to conclusively show benefits of control
- run-to-run control results in performance improvement
  with minimal capital cost
- no need to do real-time control here