PERGAMON

International Journal of Heat and Mass Transfer 44 (2001) 2053–2065

International Journal of HEAT and MASS TRANSFER

www.elsevier.com/locate/ijhmt

Recursive solution of an inverse heat transfer problem in rapid thermal processing systems

H.M. Park *, W.S. Jung

Department of Chemical Engineering, Sogang University, Shinsoo-Dong, Mapo-Gu, Seoul, South Korea

Received 10 December 1999; received in revised form 19 July 2000

Abstract

Rapid thermal processing (RTP) is a new technique for performing various wafer fabrication operations such as annealing, oxidation and chemical vapor deposition in a single chamber. The success of the RTP depends on the precise control of wafer temperature by adjusting wall heat flux. In the present investigation, an efficient recursive method is developed to solve the inverse heat transfer problem of estimating the wall heat flux on the wafer from the measurements of wafer temperature. The present recursive method is based on the Kalman filtering technique and the Karhunen–Loève Galerkin procedure. Although the direct implementation of the Kalman filter on heat conduction equation is never feasible due to the tremendous requirement of computer time and memory, a practical method of recursive estimation is devised in the present investigation by reducing the partial differential equation to a minimal set of ordinary differential equations by means of the Karhunen–Loève Galerkin procedure.

1. Introduction

Rapid thermal processing (RTP) is a new technique for performing various wafer fabrication operations such as annealing, oxidation and chemical vapor deposition in a single chamber [1]. In the manufacturing of integrated circuits, it is necessary to keep the temperature low to minimize the redistribution of dopants. But some processes, such as implant annealing, are not as effective at low temperature. Certain type of implant damage cannot be annealed out unless high temperatures are achieved. One possible avenue to minimizing diffusion during annealing at high temperature is to reduce the process time at high temperature. If RTP is employed, the dopant redistribution can be minimized during annealing by allowing brief time at high temperature [2]. One of the main technological hurdles that RTP must overcome is that of heating the wafers uniformly to prevent wafer warpage induced by the thermal stress [3]. Recent approach adopted in RTP design is the employment of multiple concentric circular rings of lamps that can be controlled independently to adjust the heat flux over the wafer to maintain a reasonable uniform temperature over a range of operating conditions [1]. Because temperature uniformity of the wafer depends critically on the distribution of wall heat flux, a real-time scheme of determining the wall heat flux distribution from the relatively easier measurements of wafer temperature has an important role in the appropriate operation of the RTP systems. With such a scheme available, one can easily manipulate the multi-lamp system to adjust the wall heat flux of the wafer such that the spatial temperature uniformity is achieved while prespecified trajectories of the wafer temperature are tracked.

In the present paper, we suggest an efficient method of determining wall heat flux from the measurements of wafer temperature. It is a typical inverse heat transfer problem. The solution of the inverse problem is not straightforward due to their ill-posedness; small perturbations in the observed functions may result into large changes in the corresponding solutions. The ill-posedness requires special numerical techniques to stabilize

* Corresponding author. Tel.: +82-2-705-8482.
E-mail address: hmpark@ccs.sogang.ac.kr (H.M. Park).

0017-9310/01/$ - see front matter
PII: S0017-9310(00)00258-1

Nomenclature

$\boldsymbol{A}$	Jacobian matrix defined in Eqs. (30)–(32)
a_i	spectral coefficient premultiplying the ith empirical eigenfunction
B_{nj}	matrix defined in Eq. (25)
$\boldsymbol{C}$	matrix defined in Eq. (34)
F_{i-j}	view factor between the ith and the jth zones
H_{ji}	matrix defined in Eq. (20)
$\boldsymbol{H}_o$	radiative flux from lamps to the upper surface of the wafer
h_w, h'_w, h''_w	heat transfer coefficients
$\boldsymbol{I}$	identity matrix
$\boldsymbol{K}(x,x')$	two-point correlation function of the Karhunen–Loève decomposition
M	number of temporal shape functions in the discretization of heat flux
M_j	vector defined in Eq. (18)
N	number of spatial shape functions in the discretization of heat flux
N_{ij}	matrix defined in Eq. (21)
NT	number of empirical eigenfunctions employed
$\boldsymbol{P}$	error covariance matrix
P_{ij}	matrix defined in Eq. (19)
P_k	total radiation from the kth lamp
$\boldsymbol{Q}$	measurement error covariance matrix
$q(r,t)$	heat flux at the wafer wall
Q_i	radiosity
R	radius of the wafer
$\boldsymbol{R}^{-1}$	model error covariance matrix
T	temperature field
$\boldsymbol{W}$	view factor matrix between lamps and upper surface of the wafer
$\boldsymbol{y}$	measurement vector (cf. Eq. (33))
Z	thickness of the wafer

Greek symbols

α_{mn}	coefficient in the discretization of heat flux (Eq. (15))
δ_{ij}	Kronecker delta
ϵ_i	emissivity
$\boldsymbol{\eta}(t)$	Gaussian white noise in the measurement
θ	deviation temperature (Eq. (1))
λ_i	the ith eigenvalue
$\boldsymbol{\xi}(t)$	Gaussian white noise in the model
ρ_i	reflectivity
σ	Stefan–Boltzmann constant
ϕ_i	the ith empirical eigenfunction
$\varphi_k^{(n)}$	the kth empirical eigenfunction for the nth set of snapshots
$\Psi_m(t)$	temporal shape function
$\Psi_n(r)$	spatial shape function
Ω	system domain

Superscripts

$*$	measurement
T	transpose

the results of calculations. Commonly adopted techniques for this purpose are the least-square methods modified by the addition of regularization terms that impose additional restrictions on admissible solutions [4] and the conjugate gradient method where the regularization is inherently built in the iterative procedure [5]. These algorithms are iterative ones and therefore require repeated computation of governing equations. Also they require a complete data base before computation begins and thus are non-recursive. However, if the scheme is to be employed in the RTP systems, it must be a recursive one where initial a priori estimation is continually updated based on the current measurements. In the present investigation, we employ the Kalman filtering technique that is a typical recursive estimation method consisting of repeatedly updating the estimates and a covariance matrix to indicate the reliability of the estimates [6–8]. But the direct implementation of the Kalman filtering technique for the solution of multi-dimensional inverse heat conduction problems is never practical due to the tremendous requirement of computer time and memory. For a two-dimensional heat conduction, the covariance equation is a four-dimensional unsteady partial differential equation. It becomes a six-dimensional partial differential equation if one considers three-dimensional heat conduction. Therefore, one of the most important prerequisites for the successful implementation of a Kalman filter for the purpose of real-time solution of inverse heat conduction problems is the development of a reliable reduced order model that is not mathematically complicated but still predicts the system behavior with accuracy. An appropriate technique for this purpose is the Karhunen–Loève Galerkin procedure [9,10] which is a Galerkin method employing the empirical eigenfunctions of the Karhunen–Loève decomposition as basis functions.

For a while, the Karhunen–Loève decomposition had been used as a rational technique enabling a stochastic field to be represented with a minimum degree of freedom [11,12]. If the Karhunen–Loève decomposition is applied to a given stochastic field, we get a set of empirical eigenfunctions. The same stochastic field can be reproduced with a minimum degree of freedom if these empirical eigenfunctions are employed [11,12]. But recent works [9,10] have extended the applicability of the Karhunen–Loève decomposition to the analysis

of non-stationary, non-homogeneous deterministic as well as stochastic fields to allow the derivation of rigorous reduced order models that simulate the given systems almost exactly. This extension of the original Karhunen–Loève decomposition is called the Karhunen–Loève Galerkin procedure [9,10]. Through the Karhunen–Loève Galerkin procedure, one can a priori limit the function space to the smallest linear subspace that is sufficient to describe the observed phenomena and consequently reduce the heat conduction equation to a minimal set of ordinary differential equations. When the Kalman filtering technique is applied to the resulting low-dimensional dynamic model, it is found to solve the inverse heat transfer problem recursively with a meager requirement of computer time and memory.

2. Thermal modeling of the RTP system

One of the most important problems with the RTP is the maintenance of thermal uniformity in the wafer. Various RTP systems have been developed that use various chamber and heating designs to improve the thermal uniformity in RTP [2,13]. In the present investigation we consider a hypothetical axisymmetric RTP system with independently controlled concentric circular rings of lamps as depicted in Fig. 1. This is in fact a simplification of the system of Moslehi et al. [13]. We adopt this simplified system to facilitate a clearer presentation of the real-time recursive algorithm which solves the inverse heat transfer problem of estimating the heat flux distribution on the wafer from the measurements of its temperature. As depicted in Fig. 1, the upper surface of the wafer is heated by radiation from the lamps which are partitioned into several circular concentric zones. The radiation intensity of the lamps in each zone is controlled independently so that the edge of the wafer, where the heat loss is greater, may receive excess radiation. The wafer and chamber surfaces are modeled as gray and diffuse, so radiative heat transfer can be computed using view factors [14,15]. In addition to the dominant radiative heat transfer from the lamps to the upper surface of the wafer, there are additional radiative heat exchanges between the wafer surfaces (upper, lower and edge) and chamber walls. Thus, the radiative heat transfer from the lamps can indirectly influence the heat flux at the lower surface and the edge of the wafer. These radiative heat exchanges are obtained by solving the enclosure problem to be described below. Our strategy in the thermal modeling in the RTP system is to separate the radiative heat exchange governed by the enclosure problem from the heat conduction within the wafer. Then, once the heat flux distribution on the surface of the wafer is estimated from the wafer temperature measurements, the lamp power distribution is trivially determined by solving the enclosure problem. The sensor locations for the measurement of wafer temperature are indicated in Fig. 1(b) with small circles. In our model of RTP, only the dominant radiative transfer from the lamps to the upper surface of the wafer is considered explicitly, and the minor radiative exchanges between the wafer surfaces (lower and edge) and the chamber walls are taken care of by the effective heat transfer coefficient. In doing so, one can reduce the condition number of the matrix in the enclosure problem and, as a result, the inverse determination of the lamp power shall become more robust.

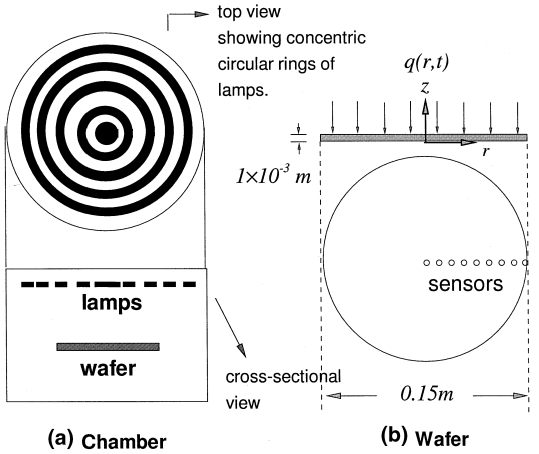

Fig. 1. The system.

Defining the deviation temperature θ as the difference between the actual temperature T' and the ambient temperature in the chamber T_a as

$$\theta = T' - T_a, \tag{1}$$

the governing equation and relevant boundary conditions for the heat conduction in the wafer under the assumption of axisymmetry are

$$\rho C_p \frac{\partial \theta}{\partial t} = k\left(\frac{1}{r}\frac{\partial}{\partial r} r \frac{\partial \theta}{\partial r} + \frac{\partial^2 \theta}{\partial z^2}\right), \tag{2}$$

$$t = 0, \quad \theta = T_i - T_a, \tag{3}$$

$$r = 0, \quad \frac{\partial \theta}{\partial r} = 0, \tag{4}$$

$$r = R, \quad k\frac{\partial \theta}{\partial r} = -h_e \theta, \tag{5}$$

$$z = 0, \quad k\frac{\partial \theta}{\partial z} = h'_w \theta, \tag{6}$$

$$z = Z, \quad k\frac{\partial\theta}{\partial z} = q(r,t) - h''_{\mathrm{w}}\theta, \tag{7}$$

where ρ is the density of the wafer, C_{p} the heat capacity, k the thermal conductivity, R the radius of the wafer, Z the thickness of the wafer, T_i the initial temperature, h_{e} the effective heat transfer coefficient at the edge, h'_{w} the effective heat transfer coefficient at the lower surface of the wafer, h''_{w} the convective heat transfer coefficient at the upper surface of the wafer, and finally $q(r,t)$ is the net radiative heat flux to the upper surface of the wafer. Since the wafer is thin enough that the axial temperature gradient can be neglected, the Eq. (2) may be integrated vertically to yield the following one-dimensional unsteady heat conduction equation governing the vertically averaged temperature field:

$$\rho C_{\mathrm{p}}\frac{\partial T}{\partial t} = \frac{k}{r}\frac{\partial}{\partial r}\left(r\frac{\partial T}{\partial r}\right) + \frac{q(r,t)}{Z} - \frac{2h_{\mathrm{w}}T}{Z}, \tag{8}$$

where $h_{\mathrm{w}} = (h'_{\mathrm{w}} + h''_{\mathrm{w}})/2$ and T is the vertical average of θ.

The relevant initial and boundary conditions for Eq. (8) are Eq. (3)–(5). The boundary conditions Eqs. (6) and (7) are incorporated into the second and the third terms in the right hand side of Eq. (8). As shown in Fig. 1(b), the domain of the system is a thin disk of radius $R = 7.5\times10^{-2}$ m and the thickness $Z = 10^{-3}$ m with the physical properties of $\rho = 2330$ kg/m^3, $C_{\mathrm{p}} = 703$ J/kg K, $k = 153$ W/m K, and $h_{\mathrm{e}} = 22.8$ W/m^2 K. The heat transfer coefficient h_{w} is assumed to vary radially according to the following empirical formula [16]:

$$h_{\mathrm{w}} = h_i + (h_o - h_i)\left(\frac{r}{R}\right)^4, \tag{9}$$

where $h_i = 28.4$ W/m^2 K and $h_o = 45.6$ W/m^2 K. According to Eq. (9), the convective and radiative cooling from the wafer is larger near the edge than at the center.

By solving the inverse heat conduction problem, the heat flux $q(r,t)$ in Eq. (7) or Eq. (8) is determined from the temperature measurements at selected locations at the upper surface of the wafer. Once $q(r,t)$ is secured, the temperature distribution in the wafer is easily obtained by solving Eq. (8). Now, we are going to explain how to pose an enclosure problem that exploits $q(r,t)$ and the wafer temperature to yield the distribution of lamp powers that have caused the observed temperature distribution on the wafer surface.

In the enclosure problem [14,15], the wafer surfaces and the chamber walls are divided into a finite number of zones and the radiative exchanges among these zones and the lamps are determined by solving a set of simultaneous linear algebraic equations. It is assumed that each zone has uniform temperature, heat flux, radiosity and radiative properties which are also isotropic and independent of frequency. The surfaces are also assumed to be opaque, diffuse emitters and diffuse reflectors. Then the enclosure problem may be posed as follows. Denoting the zones in the upper surface of the wafer with the index, $i = 1,2,\ldots,M$, the remaining zones including the lower surface of the wafer and chamber wall with the index, $i = M+1, M+2,\ldots,N$ and the lamp zones with $i = N+1, N+2,\ldots,N+L$, we have the following set of equations:

For $i = 1,2,\ldots,M$ (*upper surface of the wafer*)

$$\sum_{j=1}^{N}[\delta_{ij} - \rho_i F_{i-j}]Q_j - \rho_i H_{oi} = \epsilon_i\sigma T_i^4, \tag{10}$$

$$\sum_{j=1}^{N}[\delta_{ij} - F_{i-j}]Q_j - H_{oi} = q_i. \tag{11}$$

For $i = M+1, M+2,\ldots,N$ (*the edge of the wafer, the lower surface of the wafer and the chamber wall*)

$$\sum_{j=1}^{N}[\delta_{ij} - \rho_i F_{i-j}]Q_j = \epsilon_i\sigma T_i^4. \tag{12}$$

For each zone i, Q_i is the radiosity, ϵ_i the emissivity, ρ_i the reflectivity, T_i the temperature, F_{i-j} the view factor between the ith zone and the jth zone, q_i the net outward heat flux from the ith zone, σ the Stefan–Boltzmann constant and H_{oi} is the radiative flux from lamps to the ith zone in the upper surface of the wafer. The radiative flux from the lamps to the upper surface of the wafer is represented by

$$H_{oi} = \sum_{k=1}^{L} W_{ik}P_k, \tag{13}$$

where P_k is the total radiation from the kth lamp and W_{ik} is the view factor from the kth lamp to the ith zone in the upper surface of the wafer. Since the temperature field in the wafer and the heat flux on the upper surface of the wafer have been obtained by solving the inverse heat conduction problem and the temperature of chamber walls is given a priori, we have thus secured q_i $(i = 1,2,\ldots,M)$ and T_i $(i = 1,2,\ldots,N)$. The set of Eqs. (10)–(12) is composed of $N+M$ linear equations with $N+M$ unknowns, i.e., Q_i $(i = 1,2,\ldots,N)$ and H_{oi} $(i = 1,2,\ldots,M)$. The inversion of the matrix to obtain H_{oi} is very cheap, since the coefficients of the matrix are constants; we only have to invert it once at the beginning and store the inverse which is to be used repeatedly later. Once H_{oi} $(i = 1,2,\ldots,M)$ is known, the strengths of the lamps P_k $(k = 1,2,\ldots,L)$ that yield the

observed temperature are obtained by solving Eq. (13) in the least-squares sense as follows:

$$\mathbf{P}^{\mathrm{T}} = (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{H}_o, \tag{14}$$

where

$$\mathbf{P}^{\mathrm{T}} = (P_1, P_2, \ldots, P_L), \mathbf{W} = [W_{ik}]$$

and

$$\mathbf{H}_o^{\mathrm{T}} = (H_{o1}, H_{o2}, \ldots, H_{oM}).$$

The above explanations may be summarized as follows. From the temperature measurements at certain locations on the wafer, the heat flux distribution on the wafer is estimated by solving the inverse heat conduction problem. Using the resulting heat flux, the enclosure problem easily determines the distribution of lamp power, which gives us a clue to the readjustment of lamp powers so that the temperature gradient in the wafer be smeared out while tracking prespecified temporal trajectories of the wafer temperature. Since the radiative exchanges determined by the enclosure problem can be treated separately without significant computational burden as explained above, we shall concentrate on the recursive solution of the inverse heat conduction problem that is one of the most important ingredients for the successful operation of the RTP systems.

3. The Karhunen–Love Galerkin procedure [9,10]

The Karhunen–Loève Galerkin procedure is a Galerkin method that employs the empirical eigenfunctions of the Karhunen–Loève decomposition [11] as trial functions. The Karhunen–Loève Galerkin procedure, which had been developed in [9,10], reduces the original partial differential equation to a low-dimensional dynamic model with a minimum degree of freedom. This technique is shown to solve inverse heat conduction problems and inverse natural convection problems efficiently [17,18]. Details of the Karhunen–Loève Galerkin procedure are well documented in the references cited [9,10,17,18].

The set of empirical eigenfunctions to be employed in the Karhunen–Loève Galerkin procedure must span the solution space of the heat conduction equation for various trajectories of the boundary heat flux $q(r,t)$ under consideration. According to the Schmidt–Hilbert theory, the empirical eigenfunctions can be expressed linearly in terms of snapshots [9,10,19]. Therefore, we have to prepare an ensemble of snapshots $\{T_n(r)\}$ that encompasses the admissible solution space of the heat conduction equation for the above trajectories of the boundary heat flux $q(r,t)$.

Thus, the snapshots of the system have been obtained in the following way. As the first step, we transform the continuous heat flux $q(r,t)$ into discrete variables as follows:

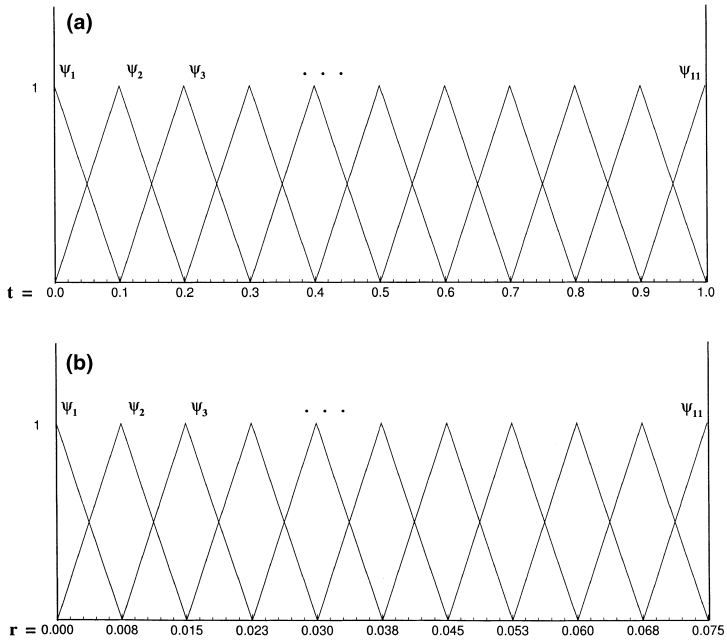

Fig. 2. Definition of shape functions. (a) Temporal shape functions. (b) Spatial shape functions.

$$q(r,t) = \sum_{m=1}^{M} \sum_{n=1}^{N} \alpha_{mn} \Psi_m(t) \Psi_n(r), \tag{15}$$

where $\Psi_m(t)$ is the mth temporal shape function employed to discretize the time variable, $\Psi_n(r)$ the nth spatial shape function employed to discretize the space variable r, M the number of temporal shape functions and N is the number of spatial shape functions employed. In the present work, we adopt $M = 11$, and $N = 11$. Fig. 2 depicts these shape functions. Next, we solve the system, Eq. (8), with $q(r,t) = \Psi_n(r)$ and record 400 transient temperature fields at an appropriate time interval until a steady state is reached. These serve as the snapshots. The Karhunen–Loève decomposition is then applied to this set of snapshots to yield empirical eigenfunctions $\{\varphi_k^{(n)}\}$, where superscript n designates the fact that these empirical eigenfunctions are obtained from the system with $q(x,t) = \Psi_n(x)$. We repeat the above procedure for $n = 1, 2, \ldots, N$ to obtain N sets of empirical eigenfunctions, i.e, $\{\varphi_k^{(1)}\}$, $\{\varphi_k^{(2)}\}$, ... ,$\{\varphi_k^{(N)}\}$. Each set $\{\varphi_k^{(n)}\}$ consists of 400 empirical eigenfunctions. Finally, we choose 10 dominant eigenfunctions from each of these N sets to make an ensemble of $(10 \times N)$ snapshots. To this set of $(10 \times N)$ snapshots, we apply the Karhunen–Loève decomposition again to obtain the final set of empirical eigenfunctions $\{\phi_k\}$ to be employed in the construction of the low-dimensional dynamic model. Fig. 3(a)–(d) show the first, second, third, and fourth empirical eigenfunctions with the corresponding normalized eigenvalues $\lambda_1 = 0.721$, $\lambda_2 = 0.199$, $\lambda_3 = 5.331 \times 10^{-2}$, and $\lambda_4 = 1.593 \times 10^{-2}$, respectively. Also shown in Fig. 4(a)–(d) are some typical eigenfunctions with smaller eigenvalues, i.e, the 22nd, the 23rd, the 24th and the 25th eigenfunctions with the corresponding normalized eigenvalues, $\lambda_{22} = 4.260 \times 10^{-10}$, $\lambda_{23} = 3.736 \times 10^{-10}$, $\lambda_{24} = 2.823 \times 10^{-10}$, and $\lambda_{25} = 2.443 \times 10^{-10}$. Figs. 3 and 4 reveal that the dominant empirical eigenfunctions represent the large scale structures of the temperature field, while the eigenfunctions with small eigenvalues represent the small scale structures.

The next step is the derivation of the low-dimensional dynamic model. We represent the temperature field $T(r,t)$ as a linear combination of the empirical eigenfunctions as follows:

$$T(r,t) = \sum_{i=1}^{NT} a_i(t) \phi_i(r), \tag{16}$$

where ϕ_i is the ith empirical eigenfunction, $a_i(t)$ the corresponding spectral coefficient and NT is the total number of the empirical eigenfunctions employed in the Karhunen–Loève Galerkin procedure. Plugging Eq. (16) into Eq. (8), applying the Galerkin principle which enforces the residual to be orthogonal to each of the trial functions and exploiting the boundary conditions, we find that

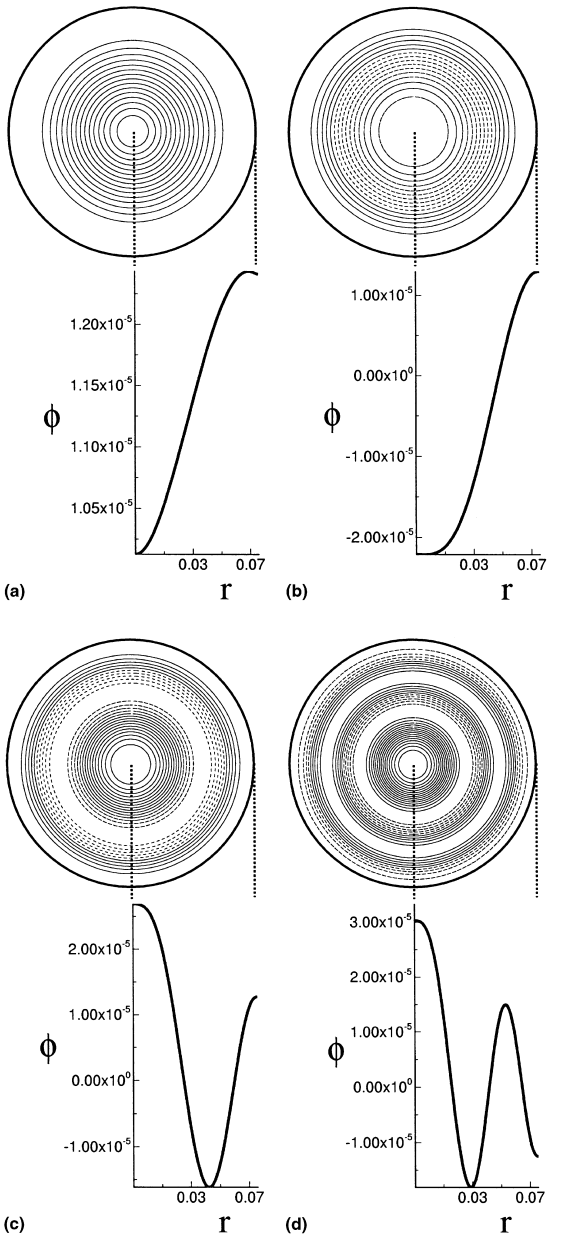

Fig. 3. Dominant empirical eigenfunctions with large eigenvalues. (a) The first eigenfunction ($\lambda_1 = 0.721$). (b) The second eigenfunction ($\lambda_2 = 0.199$). (c) The third eigenfunction ($\lambda_3 = 5.331 \times 10^{-2}$). (d) The fourth eigenfunction ($\lambda_4 = 1.593 \times 10^{-2}$).

$$\rho C_p M_j \frac{da_j}{dt} + \sum_{i=1}^{NT} a_i P_{ij} + k \sum_{i=1}^{NT} a_i H_{ji} - \frac{1}{Z} \times \int_{\Omega} q(r,t) \phi_j \, d\Omega + \frac{2}{Z} \sum_{i=1}^{NT} a_i N_{ij} = 0, \tag{17}$$

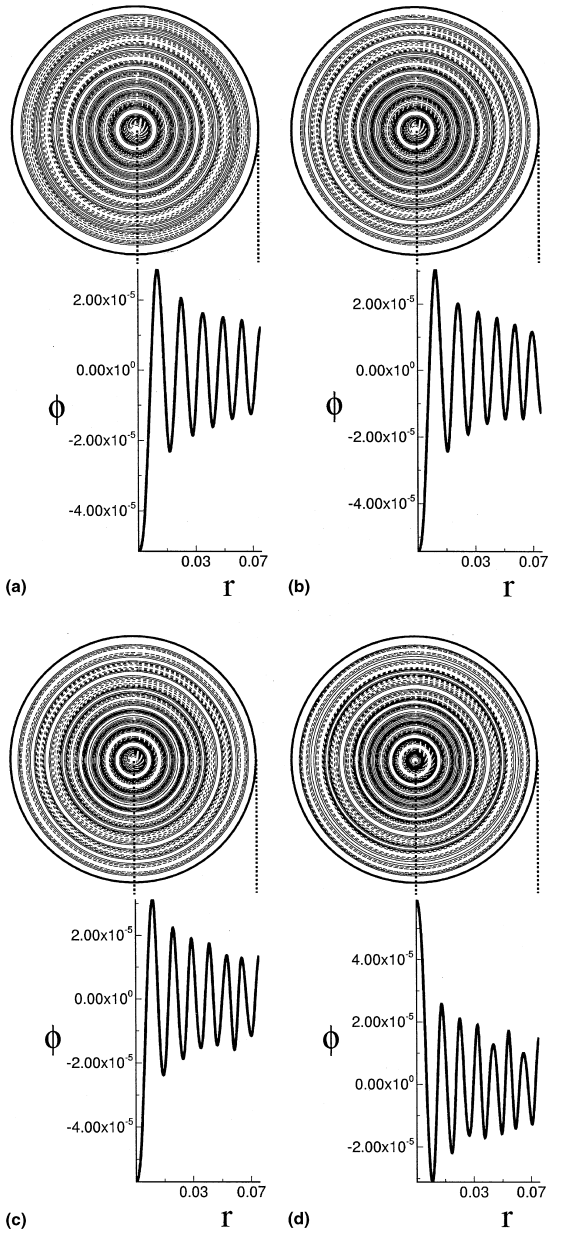

Fig. 4. Some typical empirical eigenfunctions with small eigenvalues. (a) The 22nd ($\lambda_{22} = 4.260 \times 10^{-10}$). (b) The 23rd ($\lambda_{23} = 3.736 \times 10^{-10}$). (c) The 24th ($\lambda_{24} = 2.823 \times 10^{-10}$). (d) The 25th ($\lambda_{25} = 2.443 \times 10^{-10}$).

where

$$M_j \equiv \int \phi_j^2 \, \mathrm{d}\Omega, \tag{18}$$

$$P_{ij} \equiv 2\pi R Z h_e \phi_i(R) \phi_j(R), \tag{19}$$

$$H_{ji} \equiv \int_\Omega \frac{\partial \phi_i}{\partial r} \frac{\partial \phi_j}{\partial r} \, \mathrm{d}\Omega, \tag{20}$$

$$N_{ji} \equiv \int_\Omega h_w(r) \phi_i \phi_j \, \mathrm{d}\Omega, \tag{21}$$

To corroborate the accuracy of the low-dimensional model, Eq. (17) is solved for various heat flux functions $q(r,t)$, and the resulting temperature fields are compared with those obtained by the finite difference method. Usually the error of the low-dimensional model decreases as the number of empirical eigenfunctions employed increases up to the optimal number. But further increase of number of empirical eigenfunctions beyond the optimal number does not always improve the accuracy because the empirical eigenfunctions with very small eigenvalues are contaminated with round-off errors. The optimal number of eigenfunctions for the set (23) is found to be 25, and the resulting relative errors are less than 10^{-4}. Thus, the low-dimensional dynamic model adopted in the sequel is constructed by using 25 empirical eigenfunctions.

4. Recursive estimation method

In this section, we derive a practical recursive algorithm that solves the inverse heat conduction problem of estimating the unknown wall heat flux of the wafer from the measurement of wafer temperature by applying the Kalman filtering technique to the low-dimensional dynamic model. At the beginning, the temporal domain $t \in (0, t_f)$ and the spatial domain $r \in (0,1)$ are divided into M and N segments, respectively, and $q(r,t)$ is approximated using the linear temporal shape functions $\Psi_m(t)$ (cf. Fig. 2(a)) and linear spatial shape functions $\Psi_n(r)$ (cf. Fig. 2(b)) as given by Eq. (15). Then α_{mn} in Eq. (15) is the value of heat flux at $t = t_m$ and $r = r_n$. For the interval $t \in (t_{m-1}, t_m)$, the heat flux function $q(r,t)$ may be written as

$$q(r,t) = \sum_{n=1}^{N} \alpha_{m-1,n} \Psi_{m-1}(t) \Psi_n(r) + \sum_{n=1}^{N} \alpha_{m,n} \Psi_m(t) \Psi_n(r), \tag{22}$$

where $\alpha_{m-1,n}$ $(n = 1, 2, \ldots, N)$ have been estimated during the previous time segment $t \in (t_{m-2}, t_{m-1})$ and α_{mn} $(n = 1, 2, \ldots, N)$ are the N parameters to be estimated based on the temperature measurements during the present time segment $t \in (t_{m-1}, t_m)$. Now, the low-dimensional dynamic model, Eq. (17), may be rewritten as

$$\frac{\mathrm{d}a_j}{\mathrm{d}t} = -\frac{1}{\rho C_p M_j} \sum_{i=1}^{NT} a_i P_{ij} - \frac{k}{\rho C_p M_j} \sum_{i=1}^{NT} a_i H_{ij} + \frac{1}{\rho C_p M_j} \frac{1}{\Delta z} \sum_{i=1}^{N} a_{NT+n} \Psi_m(t) B_{nj}$$

$$-\frac{1}{\rho C_p M_j}\frac{2}{\Delta z}\sum_{i=1}^{NT} a_i N_{ij}+\frac{1}{\rho C_p M_j}\frac{1}{\Delta z}$$

$$\sum_{n=1}^{N}\alpha_{m-1,n}\Psi_{m-1}(t)B_{nj}\quad (j=1,2,\ldots,NT), \tag{23}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}a_{NT+n}=0\quad (n=1,2,\ldots,N), \tag{24}$$

where

$$B_{nj}=\int_{\Omega}\Psi_n(r)\phi_j(r)\,\mathrm{d}\Omega \tag{25}$$

and new variables a_{NT+n} are assigned to the parameters α_{mn} to be estimated, i.e.,

$$a_{NT+n}=\alpha_{mn}\quad (n=1,2,\ldots,N). \tag{26}$$

Since α_{mn} $(n=1,2,\ldots,N)$ are constants during $t\in(t_{m-1},t_m)$, Eq. (24) follows. The temperature measurements at MO locations may be represented as follows:

$$T^*(r_m,t)=\sum_{j=1}^{NT}a_j(t)\phi_j(r_m)\quad (m=1,2,\ldots,MO). \tag{27}$$

Eqs. (23), (24) and (27) may be summarized in the following standard form:

$$\frac{\mathrm{d}\boldsymbol{a}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{\xi}(t), \tag{28}$$

$$\boldsymbol{y}=\boldsymbol{C}\boldsymbol{a}+\boldsymbol{\eta}(t), \tag{29}$$

where $\boldsymbol{A}$ is a $(NT+N,NT+N)$ matrix with its elements defined by

$$A_{ji}=-\frac{1}{\rho C_p M_j}P_{ij}-\frac{k}{\rho C_p M_j}H_{ij}-\frac{1}{\rho C_p M_j}\frac{2}{\Delta z}N_{ij}$$

$$(j=1,2,\ldots,NT;\ i=1,2,\ldots,NT), \tag{30}$$

$$A_{ji}=\frac{1}{\rho C_p M_j\Delta z}\Psi_m(t)B_{i-NT,j}\quad (j=1,2,\ldots,NT;$$

$$i=NT+1,NT+2,\ldots,NT+N), \tag{31}$$

$$A_{ji}=0\ (j=NT+1,NT+2,\ldots,NT+N;$$

$$i=1,2,\ldots,NT+N), \tag{32}$$

and

$$\boldsymbol{a}=\begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_{NT}\\ \hline a_{NT+1}\\ a_{NT+2}\\ \vdots\\ a_{NT+N}\end{bmatrix},\ \boldsymbol{b}=\begin{bmatrix}\Psi_{m-1}(t)/\rho C_p M_1\Delta z\sum_{n=1}^{N}\alpha_{m-1,n}B_{n1}\\ \vdots\\ \Psi_{m-1}(t)/\rho C_p M_j\Delta z\sum_{n=1}^{N}\alpha_{m-1,n}B_{nj}\\ \vdots\\ \Psi_{m-1}(t)/\rho C_p M_{NT}\Delta z\sum_{n=1}^{N}\alpha_{m-1,n}B_{n,NT}\\ \hline \boldsymbol{O}^{I}\end{bmatrix},$$

$$\boldsymbol{y}=\begin{bmatrix}T^*(r_1,t)\\ T^*(r_2,t)\\ \vdots\\ \vdots\\ \vdots\\ \vdots\\ T^*(r_{MO},,t)\end{bmatrix} \tag{33}$$

$$\boldsymbol{C}=\left[\begin{array}{ccc||c}\phi_1(r_1)\phi_2(r_1) & \cdots\cdots & \phi_{NT}(r_1) & \\ \phi_1(r_2)\phi_2(r_2) & \cdots\cdots & \phi_{NT}(r_2) & \boldsymbol{O}^{II}\\ & \vdots & & \\ \phi_1(r_{MO})\phi_2(r_{MO})\cdots\cdots\phi_{NT}(r_{MO}) & & & \end{array}\right] \tag{34}$$

In the above equations, $\boldsymbol{O}^{I}$ is a zero vector of length N, $\boldsymbol{O}^{II}$ a zero matrix of size (MO,N), $\boldsymbol{\xi}(t)$ the Gaussian white modeling noise and $\boldsymbol{\eta}(t)$ is the Gaussian white measurement noise. Using Eqs. (28) and (29), the performance function for the identification of $q(r,t)$ is expressed as follows:

$$J=\frac{1}{2}[\boldsymbol{a}(0)-\boldsymbol{a}_0]^{\mathrm{T}}\boldsymbol{P}_0^{-1}[\boldsymbol{a}(0)-\boldsymbol{a}_0]$$
$$+\frac{1}{2}\int_0^{t_f}\left\{\left[\dot{\boldsymbol{a}}-\boldsymbol{A}\boldsymbol{a}-\boldsymbol{b}\right]^{\mathrm{T}}\boldsymbol{R}\left[\dot{\boldsymbol{a}}-\boldsymbol{A}\boldsymbol{a}-\boldsymbol{b}\right]\right\}\mathrm{d}t$$
$$+\frac{1}{2}\int_0^{t_f}\left\{[\boldsymbol{y}-\boldsymbol{C}\boldsymbol{a}]^{\mathrm{T}}\boldsymbol{Q}[\boldsymbol{y}-\boldsymbol{C}\boldsymbol{a}]\right\}\mathrm{d}t, \tag{35}$$

where the weighting matrices $\boldsymbol{P}_0^{-1}$, $\boldsymbol{R}$ and $\boldsymbol{Q}$ can be chosen to reflect the errors in the initial estimate, the model and the measurement device. Employing the standard procedure [7], we can derive the recursive estimation equation for the heat flux function $q(r,t)$ as follows. The performance function, Eq. (35), is minimized using a variational method under the constraint given by Eq. (28) to yield a two-point boundary value problem. Applying the Riccati transformation to the resulting two-point boundary value

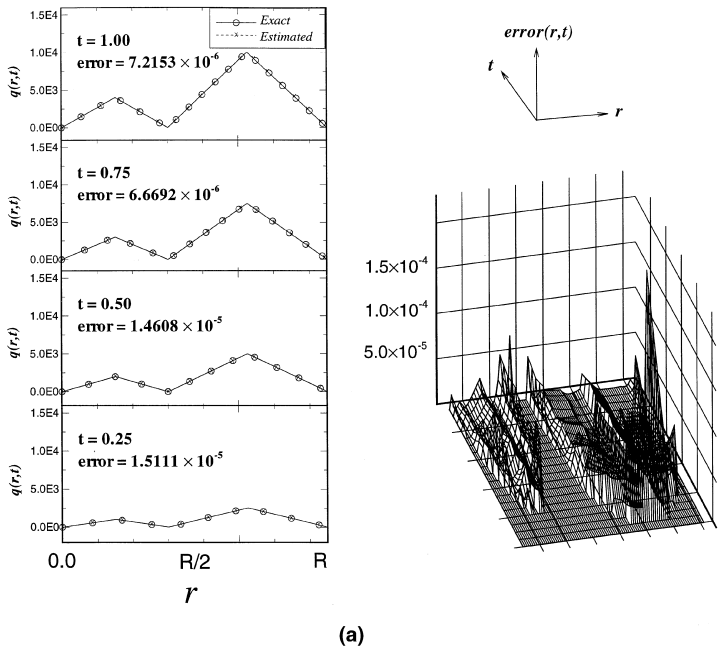

Fig. 5. The estimated profiles of wall heat flux. (a) $q(r,t)$ for the case A (Eq. 38). (b) $q(r,t)$ for the case B (Eq. 39).

problem, the following set of Kalman filtering equations is derived.

$$\frac{\mathrm{d}\boldsymbol{a}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{a} + \boldsymbol{b} + \boldsymbol{P}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}[\boldsymbol{y} - \boldsymbol{C}\boldsymbol{a}], \tag{36}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{R}^{-1} - \boldsymbol{P}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{C}\boldsymbol{P}, \tag{37}$$

where $\boldsymbol{P}$ is the error covariance matrix. Because the length of the vector $\boldsymbol{a}$ is $(NT+N)$, the covariance matrix $\boldsymbol{P}$ is of the size $(NT+N, NT+N)$ and symmetric. Thus, the number of equations to be solved to obtain $\boldsymbol{P}$ is $(NT+N)(NT+N+1)/2$.

The procedure for the recursive estimation of $q(r,t)$ is as follows. At the outset we assume initial values of a_i $(i=1,2,\ldots,NT)$ and a_{NT+n} $(n=1,2,\ldots,N)$. The initial values of $\boldsymbol{Q}$ and $\boldsymbol{P}$ are assumed to be $\boldsymbol{Q} = (1./0.0025)\boldsymbol{I}$ and $\boldsymbol{P} = P_{\mathrm{init}}\boldsymbol{I}$, where P_{init} is the largest number permissible without causing numerical difficulties. The model error covariance $\boldsymbol{R}^{-1}$ is neglected in the present computation. Solving Eqs. (36) and (37) during the first time segment, $t \in (t_1, t_2)$, we obtain α_{2n} $(n = 1,2,\ldots,N)$. During the next time segment, $t \in (t_2, t_3)$, we solve Eqs. (36) and (37) to find α_{3n} using the results of previous time segment as initial conditions except setting $P(NT+n, NT+n) = P_{\mathrm{init}}$ $(n=1,2,\ldots,N)$ at $t = t_2$. The above procedure is repeated until the final time $t = t_{\mathrm{f}}$ is reached.

5. Results

To assess the efficiency and accuracy of the present recursive algorithm of solving the inverse heat conduc-

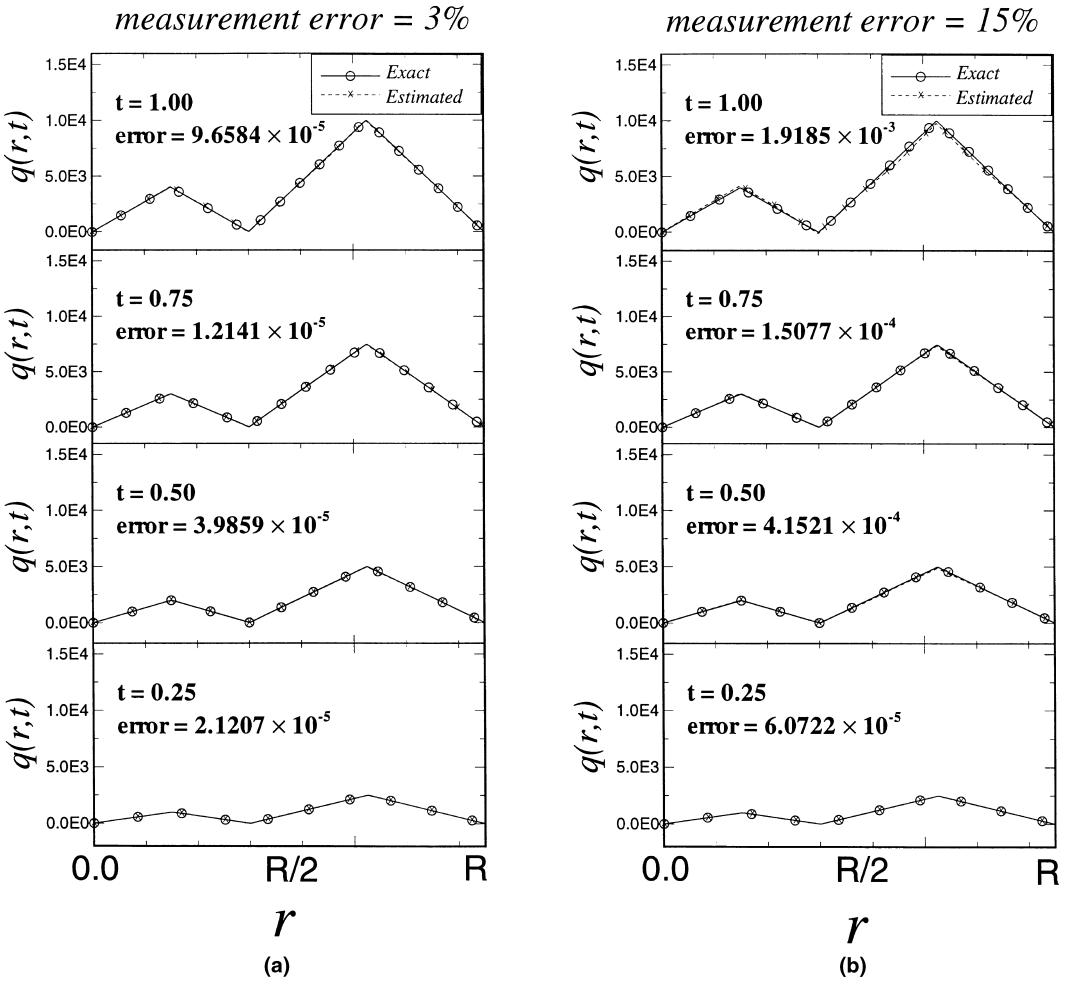

Fig. 6. Estimated $q(r,t)$ from the corrupted experimental measurements. (a) 3% relative measurement error. (b) 15% relative measurement error.

tion problem, we consider the following two heat flux functions imposed at the upper surface of the wafer:

$$\text{(case A)} \quad q(r,t) = F(r)G(t),$$
$$\text{where } G(t) = 1.0 \times 10^4 t$$
$$\text{and } F(r) = \frac{0.2}{R} r \quad \text{for } 0 \leqslant r \leqslant 0.2R, \tag{38}$$
$$= -\frac{0.2}{R}(r - 0.4R) \quad \text{for } 0.2R \leqslant r \leqslant 0.4R,$$
$$= \frac{1}{0.3R}(r - 0.4R) \quad \text{for } 0.4R \leqslant r \leqslant 0.7R,$$
$$= -\frac{1}{0.3R}(r - R) \quad \text{for } 0.7R \leqslant r \leqslant R,$$

$$\text{(case B)} \quad q(r,t) = F(r)G(t),$$
$$\text{where } G(t) = 1.0 \times 10^4 t$$
$$\text{and } F(r) = 0.0 \quad \text{for } 0 \leqslant r \leqslant 0.2R, \tag{39}$$
$$= \sin\left(\frac{5\pi(r - 0.2R)}{8R}\right) \quad \text{for } 0.2R \leqslant r \leqslant R.$$

Temperature measurements are taken with 31 sensors that are evenly distributed in the radial direction. Eq. (8) is solved with the above values of $q(r,t)$ by using a finite difference method and we adopt these numerical solutions at the measurement locations as experimental measurements after adding small random noises that are Gaussian distributed. The estimation error is given by the following equation:

$$\text{Error} = \frac{\|q_{\text{estimated}} - q_{\text{exact}}\|_{L_2}^2}{\|q_{\text{exact}}\|_{L_2}^2}, \tag{40}$$

where $\|\cdot\|_{L_2}$ is the usual L_2-norm. For the estimated values of $a_i(t=0)$ $(i = 1, 2, \ldots, NT + N)$, exact values are adopted in all computations presented in this investigation.

First, we consider an idealized situation where the measurements are not corrupted by noise. The estimated $q(r,t)$ for the two cases (Eqs. (38) and (39)) are

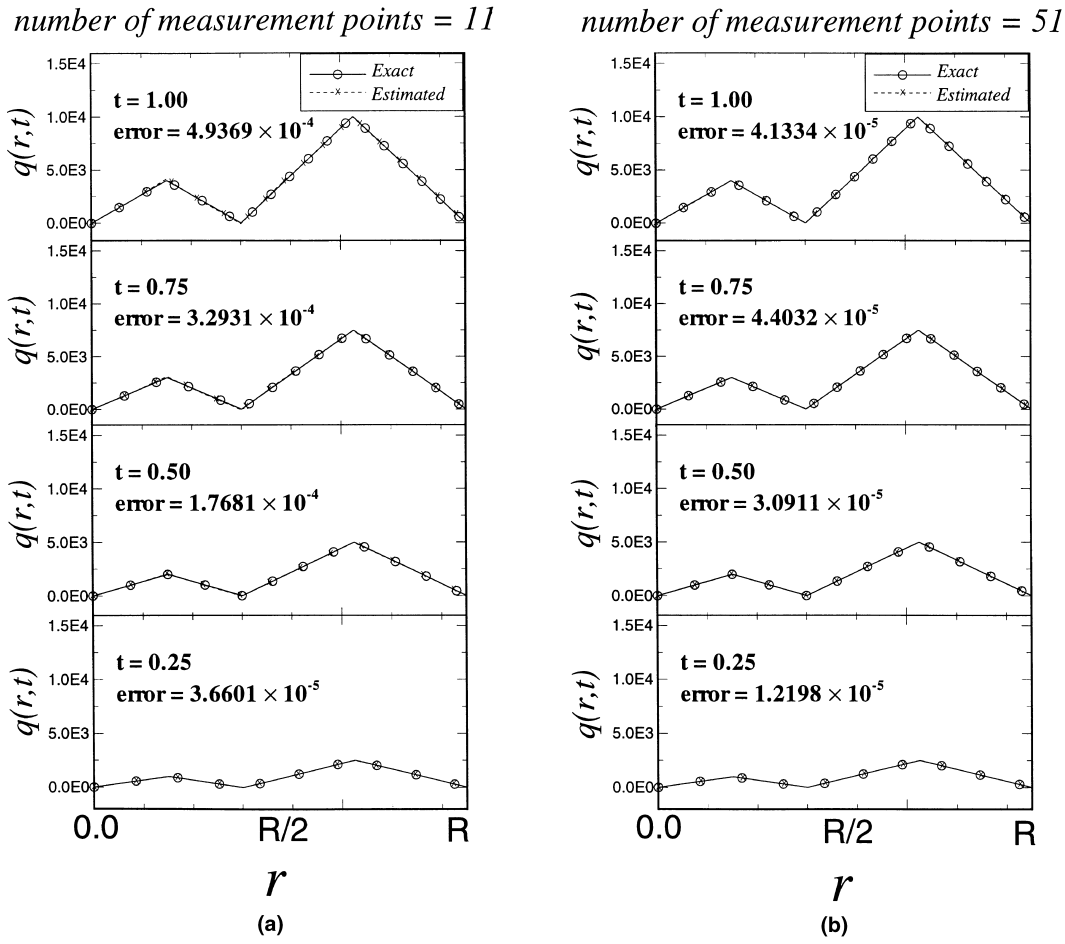

Fig. 7. Effect of number of measurement points on the accuracy of estimation. (a) 11 measurement points. (b) 51 measurement points.

shown in Figs. 5(a) and (b) with the corresponding errors. These figures show that the present method yields very accurate estimates recursively. The present method is also computationally attractive since it requires solving only $(NT+N)$ state equation (Eq. (36)) and $(NT+N)\ (NT+N+1)/2$ covariance equations (Eq. (37)) without iterations, which allow a real-time implementation of the present algorithm using a modern computer system. Next consideration is the effect of measurement noises on the accuracy of estimation. Since the measurement errors are unavoidable, a practical algorithm must be able to yield reasonably accurate estimates even with the corrupted experimental measurements. Fig. 6(a) and (b) show the estimated $q(r,t)$ when the relative measurement errors are 3% (Fig. 6(a)) and 15% (Fig. 6(b)), respectively. The estimates are still quite accurate, but it is found that the accuracy of the estimation deteriorates as the measurement noise increases if comparing these results with the estimate obtained from the exact temperature measurements (Fig. 5(a)). Now, we consider the effect of number of measurement points on the accuracy of the estimation. The default number of measurement points is 31 that are distributed evenly in the radial direction (cf. Fig. 1(b)). We employ two different sets of measurement locations; one set consists of 11 sensors and the other set 51 sensors distributed evenly in the radial direction. Fig. 7(a) shows the estimated $q(r,t)$ when employing 11 sensors and Fig. 7(b) is the corresponding result obtained with 51 sensors. Comparing these results with the default case (Fig. 5(a)), it is revealed that the accuracy of the estimation improves with the number of sensors, but the increase of sensor number beyond 31 does not improve the accuracy of the estimation.

6. Conclusion

In RTP of semiconductor wafer, precise control of wafer temperature is required throughout the process cycle to minimize dopant redistribution as well as wafer warpage. The spatial uniformity of temperature necessary during heat-up or cool-down of the wafer is ensured by an appropriate adjustment of lamp powers. Since the lamp power distribution is related with the heat flux on the wafer surface by the enclosure problem, it is important to devise a method of determining the wall heat flux on the wafer from the temperature measurements at certain locations on the wafer surface in real-time. In the present investigation, a practical recursive algorithm is devised to solve this inverse heat conduction problem by exploiting the Karhunen–Loève Galerkin procedure and the Kalman filtering technique. The Karhunen–Loève Galerkin procedure reduces the heat conduction equation to a minimal set of ordinary differential equations, and by applying the Kalman filtering technique to these ordinary differential equations, a set of equations for the recursive estimation of the unknown wall heat flux is obtained. This method is shown to yield accurate estimates recursively at a decent requirement of computer time.

References

[1] S. Norman, C. Schaper, S. Boyd, Improvement of temperature uniformity in rapid thermal processing systems using multivariable control, in: Mater. Res. Soc. Proc. vol. 224: Rapid Thermal and Integrated Processing Material Research Society, 1991.
[2] S.A. Campbell, The Science and Engineering of Microelectronic Fabrication, Oxford University Press, Oxford, 1996.
[3] F. Roozeboom, N. Parekh, Rapid thermal processing system: a review with emphasis on temperature control, J. Vac. Sci. Technol. B 8 (1990) 1249.
[4] J.V. Beck, B. Blackwell, C.R. St-Clair Jr., Inverse Heat Conduction: Ill-Posed Problems, Wiley-Interscience, New York, 1985.
[5] Y. Jarney, M.N. Özisik, J.P. Bardon, A general optimization method using adjoint equation for solving multidimensional inverse heat conduction, Int. J. Heat Mass Transfer 34 (1991) 2911–2919.
[6] R.E. Kalman, R.S. Bucy, New results in linear filtering and prediction theory, ASME J. Basic Eng. D 83 (1961) 95–108.
[7] A.H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.
[8] F. Scarpa, G. Milano, Kalman smoothing technique applied to the inverse heat conduction problem, Numer. Heat Transfer B 28 (1995) 79–96.
[9] H.M. Park, D.H. Cho, Low dimensional modeling of flow reactors, Int. J. Heat Mass Transfer 39 (1996) 3311–3323.
[10] H.M. Park, M.W. Lee, An efficient method of solving the Navier–Stokes equations for flow control, Int. J. Numer. Meth. Eng. 41 (1998) 1133–1151.
[11] M. Loève, Probability Theory, Van Nostrand, Princeton, NJ, 1955.
[12] N. Aubry, P. Holmes, J.L. Lumley, E. Stone, The dynamics of coherent structures in the wall region of a turbulent boundary layer, J. Fluid Mech. 192 (1988) 115–173.
[13] M.M. Moslehi, J. Kuehne, L. Yeakley, H. Velo, B. Najm, B. Doskalik, D. Yin, C.J. Davis, In-situ fabrication and process control techniques in rapid thermal processing, Mater. Sci. Sympos. Proc. 224 (1991) 143.
[14] M.N. Özisik, Radiative Transfer and Interactions with Conduction and Convection, Wiley, New York, 1973.
[15] M.F. Modest, Radiative Heat Transfer, McGraw-Hill, New York, 1993.

[16] H.A. Lord, Thermal and stress analysis of semiconductor wafers in a rapid thermal processing oven, IEEE Trans. Semiconductor Manufacturing 1 (1988) 105–114.
[17] H.M. Park, O.Y. Chung, J.H. Lee, On the solution of inverse heat transfer problem using the Karhunen–Loève Galerkin method, Int. J. Heat Mass Transfer 42 (1999) 127–142.
[18] H.M. Park, W.S. Jung, The Karhunen–Loève Galerkin method for the inverse natural convection problems, Int. J. Heat Mass Transfer 44 (2001) 155–167.
[19] R. Courant, D. Hilbert, Methods of Mathematical Physics, vol. 1, Interscience Publishers, New York, 1953.