Uncertainty Quantification in Molecular Signals Using Polynomial Chaos Expansion

Mahmoud Abbaszadeh, Giannis Moutsinas, Peter J. Thomas, and Weisi Guo, *Senior Member, IEEE*

***Abstract*—Molecular signals are abundant in engineering and biological contexts, and undergo stochastic propagation in fluid dynamic channels. The received signal is sensitive to a variety of input and channel parameter variations. Currently we do not understand how uncertainty or noise in a variety of parameters affect the received signal concentration, and nor do we have an analytical framework to tackle this challenge. In this paper we utilize Polynomial Chaos Expansion (PCE) to show that uncertainty in parameters propagates to uncertainty in the received signal. For demonstrating its applicability we consider a Turbulent Diffusion Molecular Communication (TDMC) channel and highlight which parameters affect the received signals. This can pave the way for future information theoretic insights, as well as guide experimental design.**

***Index Terms*—Molecular signals, fluid dynamics, uncertainty quantification.**

I. Introduction

MOLECULAR signals are abundant in biological signaling [1]–[3], industrial engineering (e.g., chemical catalysis [4]), nano-engineering [5], and ecosystems (e.g., pollution signals in rivers [6]). In many cases, the signals represent explicit information (e.g., pheromones [1]), and in other cases the chemical plume patterns are a proxy signal for an opaque process (e.g., chemical mixing-reaction rate [7]). Molecular signals exist in different scales, from nano-scale diffusion dynamics in cells to macro-scale flow in oceans. In all cases, identical molecular signal plumes never give rise to the same statistical channel response due to external disturbances and other uncertainties. Yet, simulating all possible permutations and considering all fluid dynamic forces is expensive and there is a need to quantify uncertainty more directly.

Manuscript received January 22, 2019; revised May 31, 2019; accepted August 12, 2019. Date of publication August 20, 2019; date of current version October 16, 2019. This work was supported in part by the U.S. Air Force Office for Scientific Research (AFOSR) under Grant FA9550-17-1-0056, in part by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/R041725/1, in part by the Alan Turing Institute through EPSRC under Grant EP/N510129/1, and in part by the Data-Centric Engineering Program funded by Lloyd's Register Foundation. The associate editor coordinating the review of this article and approving it for publication was T. Tugcu. *(Corresponding author: Weisi Guo.)*

M. Abbaszadeh, G. Moutsinas, and P. J. Thomas are with the School of Engineering, University of Warwick, Coventry CV4 7AL, U.K.

W. Guo is with the School of Engineering, University of Warwick, Coventry CV4 7AL, U.K., and also with Alan Turing Institute, London NW1 2DB, U.K. (e-mail: weisi.guo@warwick.ac.uk).

Digital Object Identifier 10.1109/TMBMC.2019.2936349

A. Uncertainty in Molecular Signal Propagation

Uncertainty can arise from noise in the input and ambient parameters. This makes deterministic models unreliable in estimating the variational behaviour of complex systems. Many engineering systems are described by complex differential equation models, which give deterministic outputs. In the case of mass diffusion based molecular signaling (and molecular communication), the classic Fick's Law yields a deterministic inverse-Gaussian form [8]. When considering more complex processes (e.g., turbulence, sheer stress) [2], [9], [10], the Reynolds-Averaged-Navier-Stokes (RANS) equations still yield deterministic solutions [11]. This means that externally triggered variations in input parameters (e.g., velocity profile of molecular signal) and the channel parameters (e.g., dynamic viscosity, diffusivity) cannot be accounted for. Monte-Carlo simulations are required to simulate variational behaviour, causing time intensive computation and lacking in direct insight about sensitivity to different parameter combinations.

B. Review of Similar Work

Monte-Carlo simulation, whilst time consuming, can offer computation convergence guarantees in the face of multiple uncertainties [12]. Often, as is the case for fluid dynamics, there are divergent solutions to the model. In weather forecasting, extreme weather represents one of the many possible outcomes and uncertainty propagation is essential. In [13], the authors use sparse initial weather data to inform the likelihood of divergent solutions forming. In probabilistic programming and numerics [14], uncertainty is cascaded through to yield posterior estimates of the solution, which is computationally expensive for simulations. Nonetheless, the aforementioned scenarios are computationally expensive and data demanding. As such, analytical methods for uncertainty propagation in fluid dynamics are useful.

Polynomial chaos expansion (PCE) is a method to determine the propagation of uncertainty in dynamical systems, when there is probabilistic uncertainty in the system parameters. PCE has been used in fluid dynamics since solving the Navier-Stokes equations is computationally expensive. Hosder *et al.* [15] employed non-intrusive PCE to add uncertainty in the input of aerodynamic parameters where the uncertainty was associated with the channel geometry and the dynamic viscosity of fluid in the laminar boundary-layer flow. For evaluation, they compared their results with Monte Carlo simulations and a good agreement existed. They show that

the advantage of PCE over Monte Carlo methods is that one can achieve the same results with a seven-order computational saving [15].

In molecular signaling, when the propagation channel is described by a stable mass diffusion equation, the uncertainty arises from Brownian motion [16]. However, when there are fluctuations in the diffusion channel from temperature and diffusivity variations, then uncertainty in the channel propagates to the receiver signal concentration [17]. As evidenced by the literature, there is a lack of uncertainty research in molecular-communication problems that are affected by complex fluid dynamic processes. Uncertainty research will improve our knowledge about noise sources (e.g., transposition noise [18]) and the information capacity. In this context PCE represents a good research methodology. It is worth noting that even in the whole area of wireless communication, there is a lack of research in uncertainty propagation [19], [20].

C. Contribution and Organization of Paper

In Section II, polynomial chaos expansion from a general point of view is introduced to quantify uncertainty in dynamical systems. In Section III, we introduce the molecular signal channel, including the RANS equation and integration with PCE. In Section IV, we present the results and discuss applications of PCE.

II. Polynomial Chaos Expansion

PCE is a method that facilitates the spectral representation of the uncertainty in physics-based and engineering problems. In this surrogate method, the output can be represented as an expansion of the input random parameters with a specific probability density function (PDF), so the uncertainties in the input parameters are reflected in the outputs [21].

A. Univariate Polynomial Chaos

Let Ξ be a random variable with known PDF *w*, and $X = \phi(\Xi)$, where ϕ is a function that is square-integrable on χ ($\chi \subset \mathbb{R}$) with *w* as a weight function (call this space L^2_w). Our goal is to approximate *X* by a polynomial series of Ξ. For this purpose, we need a family of polynomials H_n such that H_0 is not 0, and for $n \geq 0$ ($n \in \mathbb{N}$), the polynomial H_n has the order of *n* and are orthogonal with respect to *w*. Thus,

$$\langle H_n, H_m \rangle_w = \int_\chi H_m(x) H_n(x) w(x) dx = \gamma_m \delta_{mn}, \quad (1)$$

where δ_{mn} is the Kronecker delta and would be 1 if $m = n$ and 0 if $m \neq n$, and γ_m is the normalization constant which is obtained by:

$$\gamma_m \equiv \int_\chi H_m^2(x) w(x) dx. \quad (2)$$

We also assume that H_0 is normalized so that $\langle H_0, H_0 \rangle_w = 1$. Depending on the distribution of the Ξ, different sets of orthogonal polynomials should be employed to satisfy (1).

1) Legendre Polynomials: If Ξ has a uniform distribution ($\Xi \sim \mathrm{Unif}[-1, 1]$), Legendre polynomials should be used [22]. By having the first two Legendre polynomials ($H_0(x) = 1$ and $H_1(x) = x$), we can generate the others by following recursive relation [23]:

$$(m+1)H_{m+1}(x) = (2m+1)xH_m(x) - mH_{m-1}(x), \quad (3)$$

where *m* is the order of the polynomial. The generated polynomials are orthogonal if we consider $w(x) = \frac{1}{2}$ which is the PDF of $\Xi \sim \mathrm{Unif}[-1, 1]$. In this case, the normalization constant for $m \in \mathbb{N}$ is $1/(2m + 1)$.

2) Hermite Polynomials: When Ξ has a normal distribution ($\Xi \sim N(0, 1)$), then the Hermite polynomials should be employed to build the PCE [22]. By considering $H_0(x) = 1$, the recursive relation for Hermite polynomials is [23]:

$$H_{m+1}(x) = xH_m(x) - \frac{d}{dx}H_m(x). \quad (4)$$

The generated polynomials are orthogonal if we consider $w(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$ which is the PDF of $\Xi \sim N(0, 1)$. In this case, the normalization constant for $m \in \mathbb{N}$ is *m*!.

There are also other PDFs and their corresponding polynomials in the literature that can be considered based on the type of the input variables [22]. In this study, we utilize uniform and normal distributions as PDF of the uncertain input parameters in molecular signaling.

Now, the series is constructed by considering the polynomials H_n as a basis for L^2_w.

$$\phi(\Xi) = \sum_{n \geq 0} a_n H_n(\Xi). \quad (5)$$

where a_n are deterministic unknown coefficients. Because H_n is an orthogonal basis, the coefficients are calculated by projecting on each basis vector.

$$a_n = \frac{\langle \phi, H_n \rangle_w}{\langle H_n, H_n \rangle_w}. \quad (6)$$

After calculating the a_n coefficients, the statistics of the output *X* are determined spatially. Due to the orthogonality of the applied polynomials, the expectation of the *X* is estimated by:

$$E[X] \approx \alpha_0, \quad (7)$$

which is the coefficient of the zeroth order polynomial. Also, the variance is estimated in the same way.

$$\mathrm{Var}[X] \approx \sum_{n \geq 1} \alpha_n^2 \langle H_n, H_n \rangle_w. \quad (8)$$

Since one cannot simplify a product of three or more polynomials, the equations for further moments contain all possible terms of the power of the sum.

B. Example of Usage

Let $f(x) = xe^{-x} + x$ and let us assume that we need to find the root of the equation $f(x) = \Xi$, which we will denote by $R(\Xi)$. There is no simple close form for the root of equations of this form, but if Ξ is just a number, we can easily find the root using the Newton method. The situation becomes more

difficult if Ξ is a random variable. In this case $R(\Xi)$ can be considered a random variable itself. If we wanted to calculate the probability distribution of $R(\Xi)$, we would resort to Monte Carlo simulations, i.e., we would choose a realization of Ξ and we would compute $R(\Xi)$ of few tens of thousand times and we would plot the histogram of the values. In more complex problems this can be computationally expensive. Even calculating the mean and the variance of $R(\Xi)$ requires numerical integration of $R(\Xi)$.

Instead we will use PCE. Let us assume that $\Xi \sim \text{Unif}[-1, 1]$ and that $R(\Xi)$ can be approximated by a third order polynomial expansion using Legendre polynomials

$$R(\Xi) \approx \sum_{n=0}^{3} \alpha_n H_n(\Xi).$$

In order to calculate the coefficients a_n, we choose 10 realizations of Ξ and the corresponding root $R(\Xi)$. We will denote by ξ_i the realizations of Ξ. This means that we have 10 pairs $(\xi_i, R(\xi_i))$, so we can use least squares in order to find the optimal values for the coefficients a_n. We will call this **least square polynomial chaos expansion (LSPCE)**. We find $\alpha_0 \approx 0.042$, $\alpha_1 \approx 0.52$, $\alpha_2 \approx 0.086$ and $\alpha_3 \approx 0.012$.

From this we get that

$$E[R(\Xi)] = a_0 = 0.042$$

and

$$\text{var}[R(\Xi)] = \sum_{n=1}^{3} \alpha_n^2 \langle H_n, H_n \rangle = 0.091.$$

Furthermore, if we want to approximate the PDF of $R(\Xi)$, instead of using Monte Carlo, we can use the expansion to get values for $R(\Xi)$. In this example, instead of using the Newton method, we evaluate a third order polynomial. In more complicated problems, like turbulent diffusion problem, the gain in computational can be many orders of magnitude.

Finally, we can gauge the error of our method by calculating the fourth order approximation. In this example the first 3 coefficients were the same up to at 4 significant digits.

C. Parametric Univariate Polynomial Chaos

The case where *X* depends also on an independent parameter, *t*, can be treated as a straightforward generalization of the previous. In particular if $X = \phi(t, \Xi)$ with $\phi(t, \cdot) \in L^2_w$ for all *t*, then we can decompose *X* as

$$\phi(t, \Xi) = \sum_{n \geq 0} a_n(t) H_n(\Xi). \tag{9}$$

This becomes particularly useful in the case of differential equations that depend on a stochastic parameter. For example, in the case of a linear differential equation $\dot{X} = L(X) + \psi(\Xi)$, where *L* is linear and $\psi(\Xi) = \sum_{n \geq 0} b_n H_n(\Xi)$, we can substitute the sum and use the linearity of the equation to get

$$\sum_{n \geq 0} \dot{a}_n(t) H_n(\Xi) = \sum_{n \geq 0} L(a_n(t)) H_n(\Xi) + \sum_{n \geq 0} b_n H_n(\Xi). \tag{10}$$

and by projecting on each H_n we get the equations $\dot{a}_n = L(a_n) + b_n$, whose solutions are the coefficients of the expansion. Notice that all the equations are now deterministic.

In the case of non-linear differential equation, one has products of polynomial chaos expansions. This means that one gets an infinite system of deterministic differentiable equations. This method of computing the coefficients is called **Intrusive Method** because it requires to change drastically the solver [22].

D. Multivariate Polynomial Chaos

Let $X = \phi(\Xi_1, \Xi_2)$ where Ξ_1 and Ξ_2 are random variables with PDFs w_1 and w_2, respectively and $\phi(\cdot, \Xi_2) \in L^2_{w_1}$, $\phi(\Xi_1, \cdot) \in L^2_{w_2}$ for all Ξ_1 and Ξ_2. Let $H_{1,m}$ and $H_{2,m}$ be polynomial families that form an orthogonal basis in $L^2_{w_1}$ and $L^2_{w_2}$, respectively and $\langle H_{1,0}, H_{1,0} \rangle_{w_1} = \langle H_{2,0}, H_{2,0} \rangle_{w_2} = 1$. Then we can decompose *X* as the deterministic part (coefficients) and stochastic terms.

$$\phi(\Xi_1, \Xi_2) = \sum_{m \geq 0} \sum_{n \geq 0} a_{m,n} \, H_{1,m}(\Xi_1) H_{2,n}(\Xi_2). \tag{11}$$

One can project on the basis to get the coefficients, i.e.,

$$a_{m,n} = \frac{\left\langle \langle \phi, H_{1,m} \rangle_{w_1}, H_{2,n} \right\rangle_{w_2}}{\langle H_{1,m}, H_{1,m} \rangle_{w_1} \langle H_{2,n}, H_{2,n} \rangle_{w_2}}. \tag{12}$$

Similarly to the univariate case the relations for the first two moments are relatively simple. For the expectation one has $E[X] \approx a_{0,0}$ and for the variance we have:

$$Var[X] \approx \sum_{m \geq 1} \sum_{n \geq 1} a^2_{m,n} \langle H_{1,m}, H_{1,m} \rangle_{w_1} \langle H_{2,n}, H_{2,n} \rangle_{w_2}. \tag{13}$$

This expansion can be generalized to more than 2 random variables in a straightforward way. The PDF of each random variable defines an inner product in the space $L^2_{w_i}$ and we have to choose an orthogonal basis in this space. Then one just expands *X* with respect to every family of basis functions. Similarly we can generalize this scheme to treat parametric multivariate cases. One just needs to notice that the coefficients of the expansion depend only on deterministic parameters.

E. Approximation Using Polynomial Chaos

In order to do any computation with a PCE series, we have to truncate it since it is not feasible to expand the PCE series to infinity. For this purpose, we notice that if the series converges, then the size of the coefficients go to 0 if we take the limit of any index to infinity. This means that for every convergent series, we can ignore terms of order higher than some *N*. However, for a given problem it is not trivial to find which exactly this *N* is. Usually this is done by trial and error, where we calculate more terms until the size of the new terms is smaller than the precision required.

We start by truncating the series to an arbitrary order *N*,

$$\phi_n(\Xi) = \sum_{n=0}^{N} a_n H_n(\Xi) \tag{14}$$

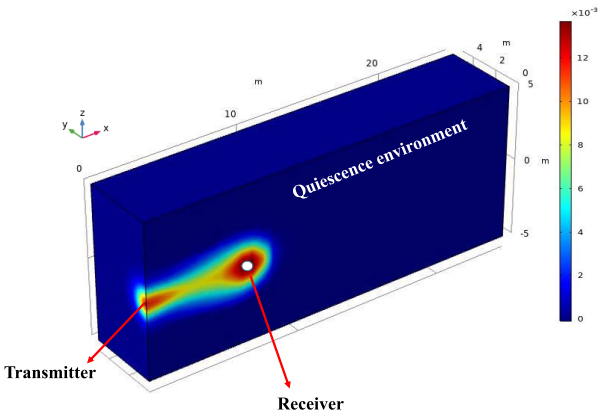


Fig. 1. Schematic of the cross-section of the system model.

and assume that this is enough for the considered precision. By using a truncated series, the intrusive method produces a finite system of differential equations which should be solved to get the coefficients.

There is also a **Non-intrusive Method** which was first introduced by [24]. In this case, we observe that (14) is linear with respect to a_n's, so we treat the simulator as a **black box**. We generate $M \geq N$ instances of the random variable Ξ, $\{\xi_1, \ldots, \xi_M\}$ and we calculate the deterministic outputs by simulator, $\{\phi(\xi_1), \ldots, \phi(\xi_M)\}$. Then for every ξ_i we have the equation:

$$\phi(\xi_i) = \sum_{n=0}^{N} a_n H_n(\xi_i). \tag{15}$$

Notice that $\phi(\xi_i)$ and $H_n(\xi_i)$ are just numbers and now we can compute the coefficients a_n by solving a linear regression. After that, we compute $\sup_{\Xi} |a_N H_N(\Xi)|$ and if it is smaller than the precision, we stop, otherwise we increase N and repeat the process.

III. System Model and Method

PCE is a surrogate approximation method for Monte Carlo simulation which has been widely used in communication area especially in molecular communication. In this study, to mimic the reality, we employed PCE in a turbulent diffusion molecular communication channel to add uncertainty in a set of parameters such as inlet velocity, initial concentration, dynamic viscosity of the ejected fluid, and the turbulent Schmidt number (or *Sc* which describes the ratio between the rates of turbulent transport of momentum and the turbulent transport of mass).

A. Channel Configuration

The system model in this study is a 3D $50 \times 10 \times 10$ m^3 water channel (see Fig. 1). The water molecules are ejected into a quiescence aqueous environment with $\mathbf{V} = u_0 i + v_0 j + w_0 k$ velocity where u_0 is taken as 4 m/s and the other two components assumed to be zero if we have an ideal injection system. The concentration of the ejected water is measured at the receiver site which is located at $40 \times r_{in}$ where r_{in} is the radius of the injector and the initial concentration of the water molecules is 4 mol/m^3. The sidewalls and the outlet are assumed to be considered far enough from the transmitter so that we can neglect their effects on the fluid flow. The properties of the water and the other system parameters are given in Table I.

TABLE I
Simulation Parameters

Variable	Value
Maximum Injection Velocity, u_0	$4\,\mathrm{m/s}$ at $t = 0$
Dynamic Viscosity of water, μ	8.9×10^{-4} Pa.s
Density of water, ρ	$1000\,\mathrm{kg/m^3}$
Transmit Concentration, c_0	$4\,\mathrm{mol/m^3}$
Pulse Width, T_0	$0.7\,\mathrm{s}$
Radius of the injector (r_{in})	$12.5\,\mathrm{cm}$
Distance Between TX and RX, $d_{Tx,Rx}$	$40 \times r_{in}$
Simulation Dimensions	$50 \times 10 \times 10\ m^3$

B. Advection-Diffusion Dynamics With RANS Equations

In order to obtain the concentration of the emitted molecules in the environment, we need to solve the advection-diffusion equation.

$$\frac{\partial c}{\partial t} = \nabla \cdot (D_\varepsilon \nabla c) - \nabla \cdot (\vec{v} c), \tag{16}$$

where c is the concentration and D_ε is the eddy diffusivity coefficient of the water molecules. c_0 is the amount of the molecules which are released into the channel at $t = 0$, and v is the velocity field of the environment flow. Generally, there are two restrictions in solving (16). First of all, $\vec{v}$ is a function of the space and time which means that in any arbitrarily location and time, the velocity components should be calculated and substituted in (16) in order to find concentration distribution. In the literature [2], this restriction has been ignored and they considered the velocity field to be constant spatially to find a closed-form relation for the concentration distribution. Secondly, the eddy diffusivity, D_ε, will be changed as the messenger molecules (MMs) go far away from the transmitter and it is not isotropic. In the literature [3], the eddy diffusivity mostly has been considered isotropic which means that the information particles in the channel can be dispersed in any directions equivalently whilst this assumption is not accurate due to the essence of the turbulent flow [25]. Based on the discussed restrictions, considering anisotropic velocity and eddy diffusivity and also, considering time-variant velocity simultaneously makes the problem complicated and finding a closed-form solution is almost impossible. In order to address the foregoing problem, the velocity distribution should be obtained and employed in (16). One of the scheme to obtain the velocity distribution is using the numerical packages to simulate the flow field and solve the **Reynolds-Average-Navier-Stokes (RANS)** equations [25]. The key characteristic of numerical packages like COMSOL Multiphysics is that they solve the RANS equations with the mass transport equation (16) simultaneously and it considers the effects of eddies

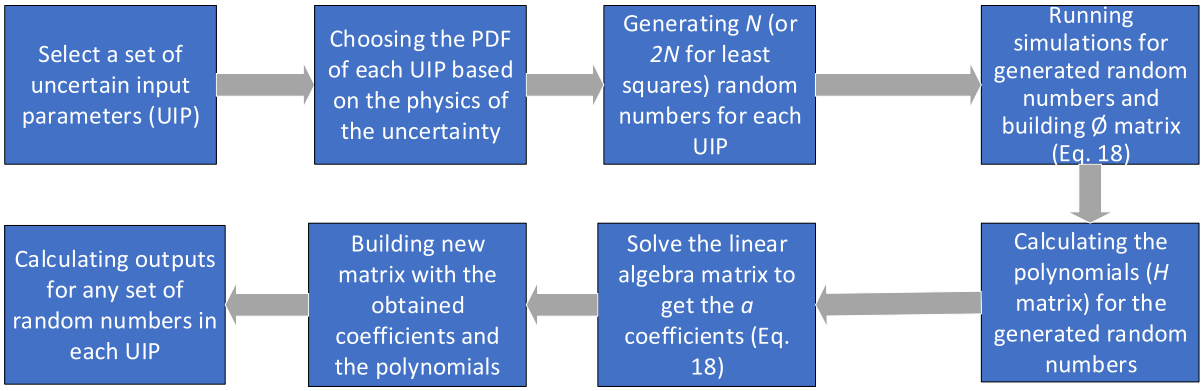


Fig. 2. Step by step procedure of building PCE.

on transporting the molecules from transmitter to receiver.

$$c\overline{u}_j \frac{\partial \overline{u}_i}{\partial x_j} = c\overline{f}_i + \frac{\partial}{\partial x_j}\left[-\overline{p}\delta_{ij} + \mu\left(\frac{\partial \overline{u}_i}{\partial x_j} + \frac{\partial \overline{u}_j}{\partial x_i}\right) - c\overline{u_i' u_j'}\right], \tag{17}$$

where c represents density or concentration which depends on a number of parameters such as pressure, velocity, and sheer stress gradients. The dynamic viscosity of the fluid is μ, and the term $c\overline{u}_j \frac{\partial \overline{u}_i}{\partial x_j}$ represents the change in mean momentum of the fluid element due to the unsteadiness in the mean flow and the convection by the mean flow. This is balanced by the mean body force $\overline{f}_i$, the isotropic stress from the pressure field $\overline{p}\delta_{ij}$, the viscous stresses, and the apparent stress $-c\overline{u_i' u_j'}$ owing to the fluctuating velocity field (Reynolds stress). Whilst there are statistical approximate solutions in the form of eddy diffusivity, general tractability is still a challenge for modeling turbulent diffusion and that is why finite-element simulation is used.

C. Non-Intrusive PCE

In practice, it is impossible to be certain about the exact values of the initial conditions in any applied problem since the input parameters are subject to uncertainties that originate from various sources (e.g., injection delay, concentration of information molecules, etc). For example, in our problem, when we say that the inlet velocity and initial concentration are $\mathbf{V} = \alpha_1 i + \alpha_2 j + \alpha_3 k$ and β mol/m^3, respectively, they are not precisely these values and they can be $\mathbf{V} = (\alpha_1 \pm \Delta\alpha_1)i + (\alpha_2 \pm \Delta\alpha_2)j + (\alpha_3 \pm \Delta\alpha_3)k$ and $\beta \pm \Delta\beta$ mol/m^3 where $\Delta\alpha_1, \Delta\alpha_2, \Delta\alpha_3$, and $\Delta\beta$ are the amount of uncertainty in each input parameters. For adding such uncertainties in our simulation, we employed the non-intrusive PCE method meaning that we did not have to make any changes in the built-in solver code (see Section II-D). For constructing the PCE, the expansion in (11) should be truncated (complete description in Section II-D). Generally, there are two ways of truncating the PCE: 1) truncating the series up to an order of interest. For example, if the order of interest is z, then for the series of (11), we have:

$$\phi(\Xi_1, \Xi_2) = \sum_{0 \le (n+m) \le z} a_{m,n}\, H_{1,m}(\Xi_1) H_{2,n}(\Xi_2). \tag{18}$$

2) We can consider all possible combinations of the order for each univariate polynomial. This method which is called **tensor product truncation**, given by:

$$\phi(\Xi_1, \Xi_2) = \sum_{0 \le n \le z} \sum_{0 \le m \le z} a_{m,n}\, H_{1,m}(\Xi_1) H_{2,n}(\Xi_2). \tag{19}$$

In this paper, we utilize the first method since it is time-efficient and does not substantially affect the accuracy [21]. If the number of the uncertain input parameters is n and the maximum order of the applied polynomial is p, then the number of the terms in the truncated PCE (or size of experimental design) would be $\binom{N=n+p}{p}$ for the first truncation method, and for the tensor product truncation it would be $N = (p+1)^n$ [21]. After generating N random numbers for each uncertain input parameter (Legendre polynomials are required if we choose a uniform distribution and Hermite polynomial for a normal distribution), the COMSOL code should be evaluated for these random numbers to obtain the left hand side of (15). Thereafter, the polynomials on the right hand side of (15) are also calculated at these points and finally, by solving a linear regression problem, the a_n coefficients will be obtained. It should be noted that all of the terms in (15) are matrices. Now, we can construct the expansion and use it as an approximation formula without running the solver for a new set of random numbers for each uncertain input parameter. The block diagram that shows each step in building the PCE is shown in Fig. 2. What is crucial in the context of the regression is the size of the experimental design or N. If we evaluate the PCE exactly at N random numbers, it does not result in a stable series and one finds oscillations in the established PCE (see Fig. 3-a1, -b1, -c1). In Fig. 3-a1, -b1, and -c1, each line represents the concentration as a function of time for each set of random numbers. For example, for the generated random number for the velocity vector ($\mathbf{V} = 3.95i + 0.023j - 0.145k$), and for the initial concentration 3.987 mol/m^3, we obtain one of the blue lines in Fig. 3-a1 and for a different set of random numbers for these uncertain parameters we then also obtain a different blue line. In total there are N lines. Fig. 3-a1,-b1, and -c1 should display a behaviour that is qualitatively similar to the simulation results (see inset of Fig. 3-a2, -b2, and -c2) to give sufficient confidence in the accuracy of the PCE. Reference to the figures does however reveal that this is not

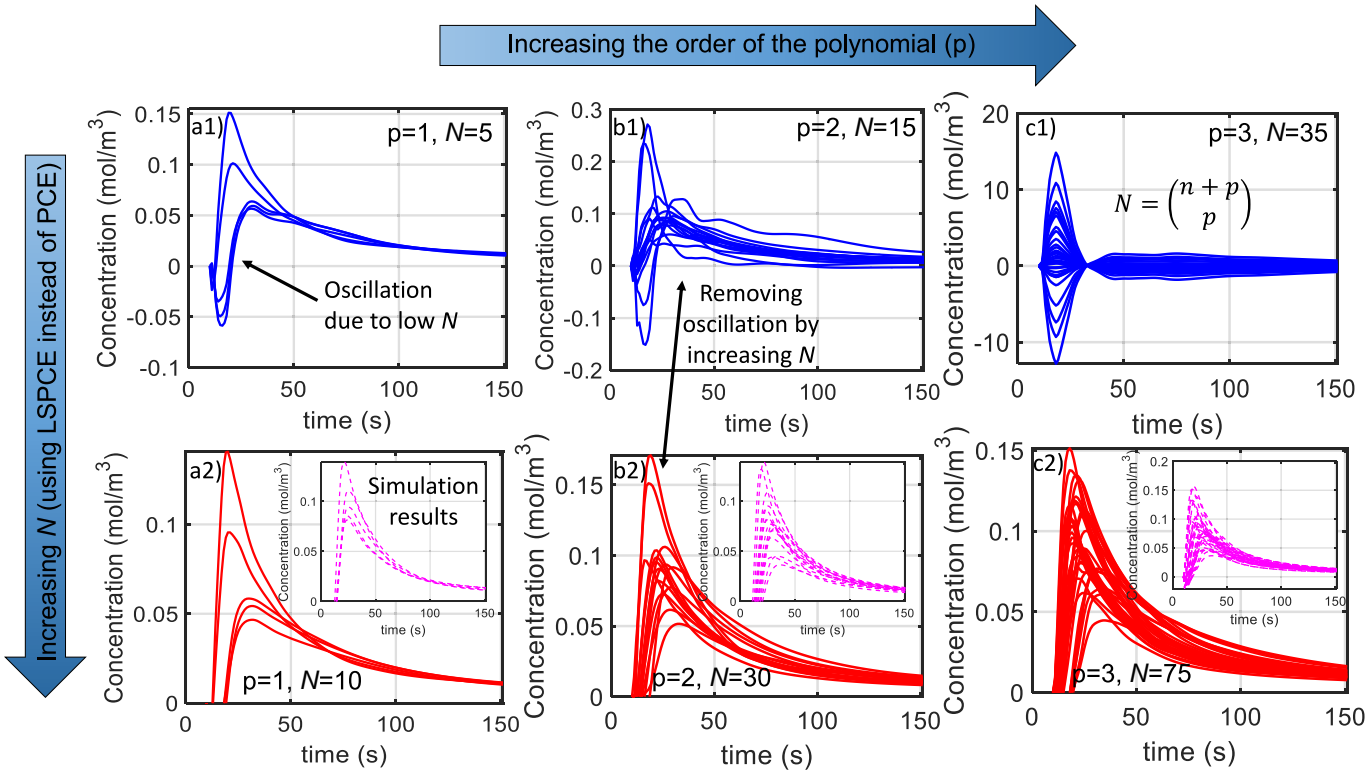


Fig. 3. Concentration vs. time profiles. From left to right shows the increasing of the polynomial order (p). First row is for the PCE and the second row is for the LSPCE. In each column, the p is fixed for better illustration of the increased N. Simulation results also is inset in the LSPCE plots for better comparisons and seeing how similar are the LSPCE and simulation results.

the case. To resolve this problem, one can choose the size of the experimental design $2 \times N$ as suggested by many authors in literature [21], [26]. After evaluating the series in $2 \times N$ random numbers, it is required to interpolate the results by means of the least-squares regression (LSPCE). This formalism removes the oscillations and yields a stabilized series (see Fig. 3-a2, -b2, -c2). It should be noted that the simulation results in Fig. 3 are Monte Carlo results with very low accuracy. For high accuracy Monte Carlo results it requires many more numerical simulations (e.g., 1000 simulation runs) which is highly time-consuming. A sample of results obtained from numerical simulations are included in Fig. 3 to illustrate the similarity with the LSPCE data.

IV. Results & Discussion

A. Convergence Test of PCE

The number of the uncertain input parameters (n) of any particular problems is fixed. Increasing the order (p) of the polynomial improves the convergence of the values of the coefficients ($a_{m,n}$). However, increasing the order of the polynomial means that a substantially higher number of simulations is required. Therefore a compromise between accuracy and required computational time is necessary.

Reference to the graphs in Fig. 3-a2, -b2, -c2 reveals that it is not possible to infer which order of p yields sufficient convergence of the PCE process. Therefore the statistics of the random outputs have to be considered. To this end the expectation and the variance of the concentration in the Fig. 3-a2, -b2, -c2 have to be calculated from (7) and (8), respectively. These two quantities are shown in Fig. 4. It can be seen that the expectation for $p = 2$ and $p = 3$ almost overlap while there is a significant difference for $p = 1$ in comparison to $p = 2$. The inset in Fig. 4-a1 displays the absolute difference of the expectations for $p = 2$, $p = 3$ and $p = 2$, $p = 1$. The figure reveals that the difference is one order of magnitude. Thus, using a polynomial of order $p = 1$ is certainly not sufficient.

The change of the variance by increasing the polynomial order is displayed in Fig. 4-a2. It can be seen that the variance between cases $p = 2$ and $p = 3$ changes by one order of magnitude whereas for $p = 1$ and $p = 2$ the change is two orders. Thus, $p = 2$ can be considered as an appropriate choice for the polynomial order. Choosing higher order polynomials substantially increases the required simulation time with only minor effects on improving the accuracy of the results.

B. Comparing Different PDFs in PCE

Depending on the input variables different types of PDFs have to be employed to obtain the most reliable results in PCE. For instance, for the problem defined in Section III the dynamic viscosity of water (μ) and the turbulent Schmidt number (Sc) are the uncertain input parameters. The dynamic viscosity varies with temperature (it is 8.9×10^{-4} Pa.s at 25°C). The uncertainty of the turbulent Schmidt number arises since it is determined experimentally. Fig. 5 displays predictions for the concentration as a function of time using normal and uniform PDF distributions for both the viscosity and the Schmidt number. The figure reveals that the curves

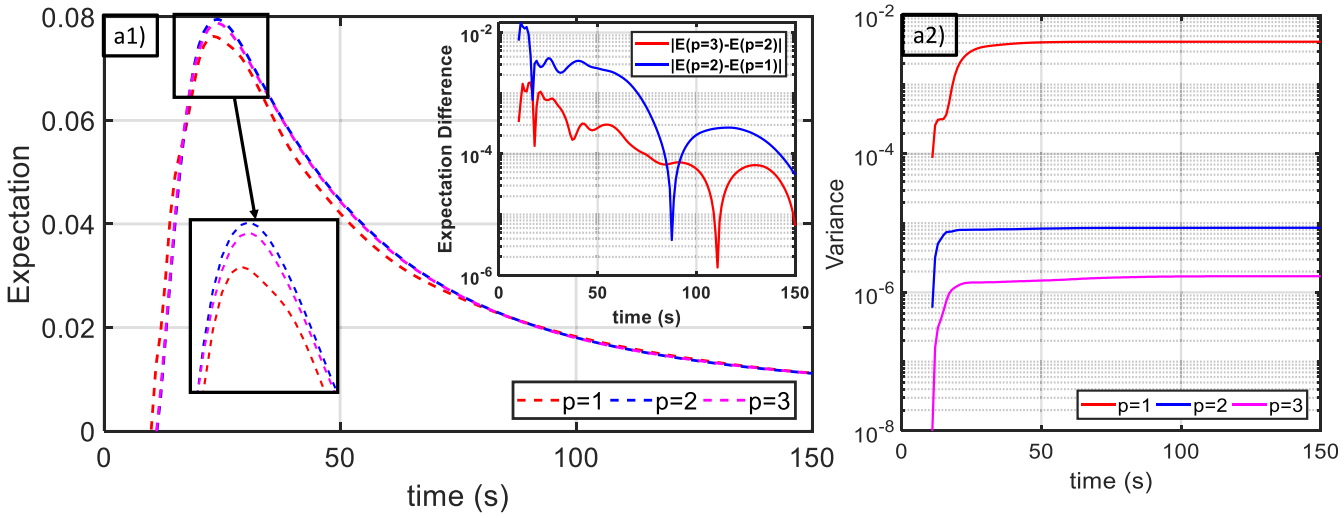


Fig. 4. a1) Expectation of LSPCE for different polynomial orders. The inset shows the difference between the expectations by increasing the order of the applied polynomials. a2) Variances of LSPCE for different polynomial orders.

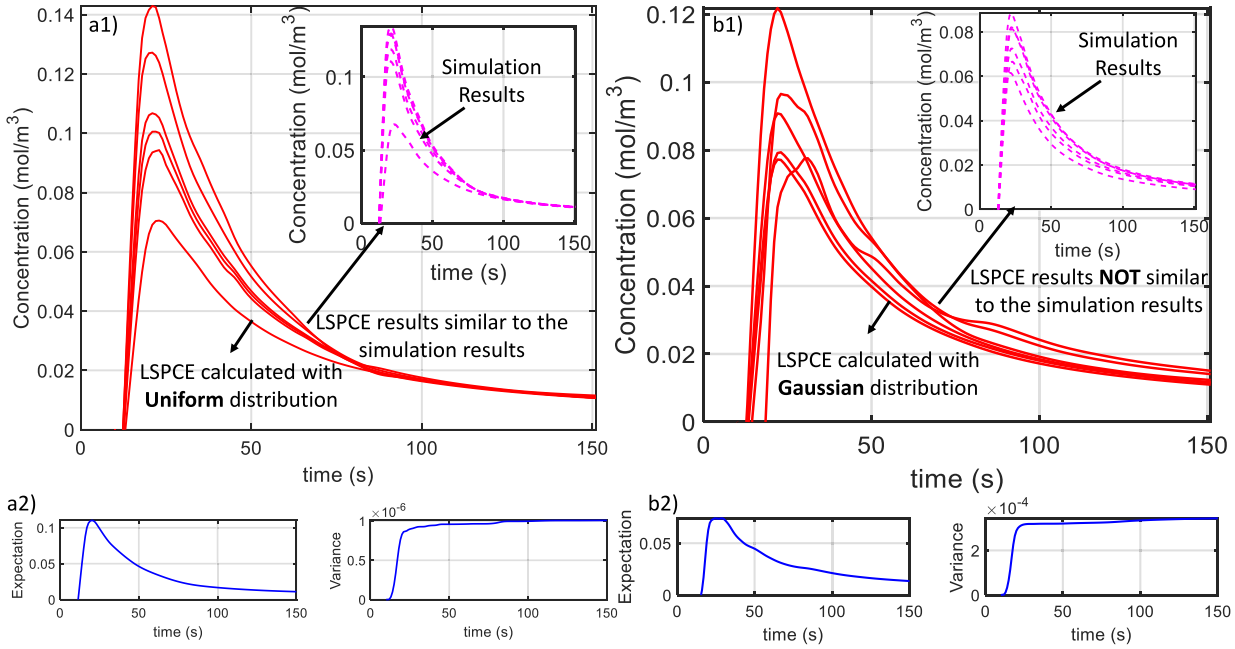
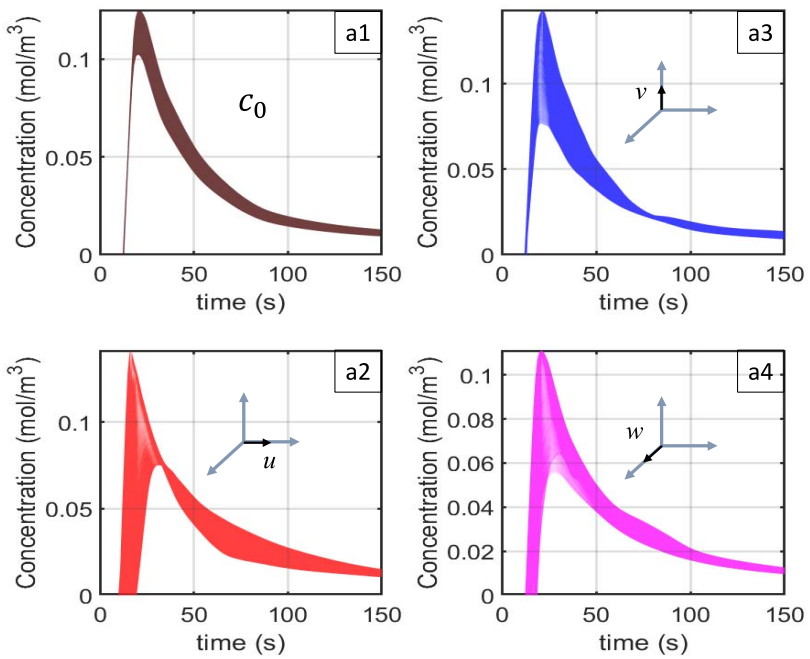


Fig. 5. a1) Concentration vs. time using uniform distribution as the PDF of the uncertain input parameters, dynamic viscosity of the water (μ) and turbulent Schmidt number(Sc), a2) The expectation and variance of the concentration corresponding to the uncertain input parameters with uniform distribution. b1) Concentration vs. time using Gaussian distribution as the PDF of the uncertain input parameters, b2) The expectation and variance of the concentration corresponding to the uncertain input parameters with Gaussian distribution.

for the concentration profiles when using a uniform PDF are smoother than those based on a normal distribution (see Fig. 5-a1 and -b1). This is a result of the uniform distribution being bounded, between -1 and 1, whilst the normal distribution is not. That is, for input parameters like the dynamic viscosity of the water which is small and only subject to small variations considering an unbounded PDF would not yield reliable results. This issue is further highlighted by the results for the expectation and variance shown in Fig. 5-a2 and -b2. The variance for a uniform PDF in Fig. 5-a2 is two orders of magnitude smaller than that for the normal PDF in Fig. 5-b2.

C. Statistical Results

1) Multiple Uncertainties: In Fig. 6 we employed the LSPCE approach to investigate the statistical properties of the TDMC channel with one of the input parameters being subject to uncertainty only. In Fig. 6-a1 the uncertain parameter is c_0, in Fig. 6-a2 to -a4 the uncertain parameters are different components of the inlet velocity ($\mathbf{V} = u_0 i + v_0 j + w_0 k$). Six simulations per parameter ($\frac{(n+p)!}{n!p!} \times 2 = \frac{(1+2)!}{2!} \times 2 = 3 \times 2 = 6$ LSPCE) were performed to construct the PCE. The expansion was then used to calculate the concentration for 10000 different cases, and the random number being generated from a uniform distribution in each case. The ratio between



a5

Parameter	Average signal peaks to average
c	3.2088
u	2.9411
v	3.2969
w	2.8714

a6

Parameter	Variance in signal peaks
c	0.0071
u	0.0188
v	0.0187
w	0.0163

Fig. 6. 10000 concentration profiles obtained by LSPCE for uncertain input parameters of (a1) c_0 (a2) u_0, (a3) v_0 (a4) w_0. (a5) Average signal peaks to average and (a6) Variance in signal peaks for all of 10000 concentration profiles for each uncertain input parameter.

the computational times required to arrive at the same result by means of 10000 3D COMSOL Monte Carlo simulations and the LSPCE simulations is of the order of 10^5.

2) Communication Performance Impact: The varying color density displayed by the plots in Fig. 6-a1-a4 moreover illustrates that the uncertainty in the input is mostly reflected in the peak of the channel response while the tail is not affected. This means that only the signal strength is subjected to be affected by the uncertainty while the inter-symbol-interference (ISI) is not.

It can also be seen that the channel response is more sensitive to the uncertainty of the inlet velocity compared to the initial concentration. This represents a crucial result in the context of experiments. It implies that geometric details of piston and cylinder of the transmitter, which govern the inlet velocity have a stronger effect on measured data than the amount of the released molecules.

Finally, the received ratio of average signal peaks to average is important from the perspective of the power amplifier linearity, which needs to convert molecular count into a received signal. The table in Fig. 6-a5 illustrates that the ratio is not significantly affected by introducing uncertainty in the initial concentration or inlet velocity. Table in Fig. 6-a6, which displays the variance in signal peaks, demonstrates that uncertainty in the initial concentration leads to more noise in the channel compared to uncertainty in the inlet velocity.

V. Conclusion & Future Work

Currently we do not understand how uncertainty or noise in a variety of parameters affect the received signal concentration, and neither do we have an analytical framework to tackle this challenge. In this paper, we utilized Polynomial Chaos Expansion (PCE) to show how uncertainty in parameters propagates to uncertainty in the received signal. The PCE method has a significant time saving compared to Monte-Carlo simulations and offers theoretical insight. The analytical results are validated using multi-physics COMSOL simulations in a Turbulent Diffusion Molecular Communication (TDMC) channel. Our uncertain parameters are initial concentration, injection velocity, dynamic viscosity of the water, and turbulent Schmidt number. We demonstrated that how the uncertainty in the aforementioned parameters propagates through the channel and can affect the received signal response.

The research conducted here in PCE and uncertainty propagation can pave the way for future information theoretic insights, as well as guide experimental design. In the future, we will focus on understanding the channel capacity as a function of uncertainty.

References

[1] T. D. Wyatt, *Pheromones and Animal Behavior: Chemical Signals and Signatures*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[2] B. D. Unluturk and I. F. Akyildiz, "An end-to-end model of plant pheromone channel for long range molecular communication," *IEEE Trans. Nanobiosci.*, vol. 16, no. 1, pp. 11–20, Jan. 2017.

[3] N. Farsad, H. B. Yilmaz, A. Eckford, C.-B. Chae, and W. Guo, "A comprehensive survey of recent advancements in molecular communication," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1887–1919, 3rd Quart., 2016.

[4] M. Icardi, G. Boccardo, D. Marchisio, T. Tosco, and R. Sethi, "Pore-scale simulation of fluid flow and solute dispersion in three-dimensional porous media," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 90, no. 1, 2014, Art. no. 013032.

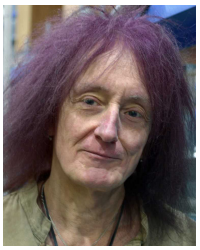
[5] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, "The Internet of bio-nano things," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 32–40, Mar. 2015.

[6] F. Sonnenwald, V. Stovin, and I. Guymer, "Configuring maximum entropy deconvolution for the identification of residence time distributions in solute transport applications," *J. Hydrol. Eng.*, vol. 19, no. 7, pp. 1413–1421, 2014.

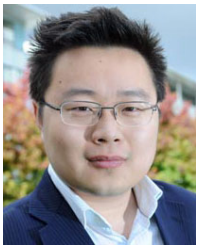
[7] A. Guida, A. W. Nienow, and M. Barigou, "Shannon entropy for local and global description of mixing by Lagrangian particle tracking," *Chem. Eng. Sci.*, vol. 65, no. 10, pp. 2865–2883, 2010.
[8] W. Guo, T. Asyhari, N. Farsad, H. Yilmaz, A. Eckford, and C. Chae, "Molecular communications: Channel model and physical layer techniques," *IEEE Wireless Commun.*, vol. 23, no. 4, pp. 120–127, Oct. 2016.
[9] E. Kennedy, P. Shakya, M. Ozmen, C. Rose, and J. Rosenstein, "Spatiotemporal information preservation in turbulent vapor plumes," *Appl. Phys. Lett.*, vol. 112, Jun. 2018, Art. no. 264103.
[10] M. Abbaszadeh, P. J. Thomas, and W. Guo, "Towards high capacity molecular communications using sequential vortex rings," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 4, no. 1, pp. 39–42, Mar. 2018.
[11] M. Abbaszadeh, H. B. Yilmaz, P. J. Thomas, and W. Guo, "Linearity of sequential molecular signals in turbulent diffusion channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2019, pp. 1–6.
[12] L. Mathelin, M. Y. Hussaini, and T. A. Zang, "Stochastic approaches to uncertainty quantification in CFD simulations," *Numer. Algorithms*, vol. 38, nos. 1–3, pp. 209–236, 2005.
[13] M. Farazmand and T. P. Sapsis, "A variational approach to probing extreme events in turbulent dynamical systems," *Sci. Adv.*, vol. 3, no. 9, 2017, Art. no. e1701533.
[14] P. Hennig, M. A. Osborne, and M. Girolami, "Probabilistic numerics and uncertainty in computations," *Proc. Roy. Soc. A*, vol. 471, no. 2179, pp. 1–17, Jun. 2015.
[15] S. Hosder, R. Walters, and R. Perez, "A non-intrusive polynomial chaos method for uncertainty propagation in CFD simulations," in *Proc. 44th AIAA Aerosp. Sci. Meeting Exhibit*, 2006, p. 891.
[16] K. V. Srinivas, A. W. Eckford, and R. S. Adve, "Molecular communication in fluid media: The additive inverse Gaussian noise channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4678–4692, Jul. 2012.
[17] S. Qiu *et al.*, "Molecular channel fading due to diffusivity fluctuations," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 676–679, Mar. 2017.
[18] W. Haselmayr, S. M. H. Aejaz, A. T. Asyhari, A. Springer, and W. Guo, "Transposition errors in diffusion-based mobile molecular communication," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 1973–1976, Sep. 2017.
[19] H. Acikgoz and R. Mittra, "Statistical analysis of electromagnetic structures and antennas using the polynomial chaos expansion method," in *Proc. IEEE 11th Eur. Conf. Antennas Propag. (EUCAP)*, 2017, pp. 798–800.
[20] F. Boeykens, H. Rogier, and L. Vallozzi, "An efficient technique based on polynomial chaos to model the uncertainty in the resonance frequency of textile antennas due to bending," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1253–1260, Mar. 2014.
[21] N. E. Owen, "A comparison of polynomial chaos and Gaussian process emulation for uncertainty quantification in computer experiments," Ph.D. dissertation, Dept. Math., Univ. Exeter, Exeter, U.K., 2017.
[22] D. Xiu and G. E. Karniadakis, "Modeling uncertainty in flow simulations via generalized polynomial chaos," *J. Comput. Phys.*, vol. 187, no. 1, pp. 137–167, 2003.
[23] T. S. Chihara, *An Introduction to Orthogonal Polynomials*. Mineola, NY, USA: Courier Corporat., 2011.
[24] R. Walters, "Towards stochastic fluid mechanics via polynomial chaos," in *Proc. 41st AIAA Aerosp. Sci. Meeting Exhibit*, Reno, NV, USA, 2003, pp. 1–13.
[25] P. J. Roberts and D. R. Webster, *Turbulent Diffusion*. Reston, VA, USA: ASCE Press, 2002.
[26] S. Hosder, R. W. Walters, and M. Balch, "Efficient sampling for non-intrusive polynomial chaos applications with multiple uncertain input variables," in *Proc. 48th AIAA/ASME/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf.*, 2007, p. 1939.

Mahmoud Abbaszadeh received the B.S. degree in mechanical engineering, heat and fluids, and the M.Sc. degree in mechanical engineering and energy conversion from the University of Kashan, Iran, in 2013 and 2015, respectively. He is currently pursuing the Ph.D. degree in molecular communication with the University of Warwick, U.K. His research interests include information capacity of turbulent diffusion channel, uncertainty quantification, and DNA data storage and communication.

Giannis Moutsinas received the degree in physics (with a major in mathematical physics) from the Aristotle University of Thessaloniki in 2008, the M.Sc. degree in mathematics from Utrecht University in 2011, and the Ph.D. degree in mathematics from the University of Warwick in 2016, studying the asymptotic behavior of dynamical systems. Currently, he is a Research Fellow with the School of Engineering, University of Warwick. His research interests include dynamical systems, complexity theory, and uncertainty quantification.



Peter J. Thomas received the Dipl.-Phys. and Dr.rer.nat. degrees in physics from the Georg-August-Universität, Göttingen, Germany, in 1988 and 1991, respectively. He was with the Max-Planck Institute of Fluid Mechanics and Applied Mechanics, Göttingen, and then the German Aerospace Centre (DLR), Göttingen. In 1992, he joined the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, U.K. In 1995, he joined the University of Warwick, where he was promoted to Professor in 2008 and where he is the Director of the Fluid Dynamics Research Centre. His current research interests include the experimental and theoretical investigations of problems in fluid dynamics and of nonlinear mechanical processes in general. He was a recipient of the International Union of Applied and Theoretical Mechanics Bureau Prize in 1996. He is a member of the European Mechanics Society.



Weisi Guo (S'07–M'11–SM'17) is an Associate Professor and a Turing Fellow. He heads the Data-Embedded-Networks Laboratory, University of Warwick. His research interests include molecular communications, wireless networks, complex networks, and social physics. He has published over 110 peer-reviewed papers, won awards from the IEEE and IET, and been a finalist for the Bell Labs Prize. His research has been widely published on the BBC, *Nature*, and *The Economist*.