# Symbol Detection and Molecule Mixture Design for Non-linear Receiver Arrays in MC

Bastian Heinlein[1,2], Sümeyye Yilmaz[1], Kaikai Zhu[1], Sebastian Lotter[1], Helene M. Loos[1,3,4], Andrea Buettner[1,4], Robert Schober[1], and Vahid Jamali[2]
[1] *Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*
[2] *Technical University of Darmstadt, Darmstadt, Germany*
[3] *Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany,*
[4] *Fraunhofer Institute for Process Engineering and Packaging IVV, Freising, Germany*

***Abstract*—Air-based molecular communication (MC) has the potential to provide the first practical applications of MC systems thanks to already existing sensor technology. However, the existing sensors behave imperfectly and non-linearly, contrary to the common assumptions in the MC literature. To overcome this mismatch, we propose a detector for a non-linear receiver array and an algorithm for constructing a molecule mixture shift keying (MMSK) modulation alphabet optimized for this receiver array. We evaluate the performance of the proposed detector and algorithm through simulations based on measurements reported for two commercial sensors.**

## I. Introduction

Air-based MC has the potential to realize disruptive applications, e.g., in environmental monitoring and environmental engineering [1]. However, while existing sensor technology could render air-based MC practically feasible in the near future, existing sensors are imperfect and often show non-linear behavior, which is only rarely considered in the MC literature [2]. One particular example of such sensors are metal-oxide semiconductor (MOS) sensors, which exhibit a power-law response to gas concentrations and are cross-reactive, i.e., they are not perfectly specific to a single gas but respond to multiple gases [3]. To account for these properties, we propose a simple, yet general detector that is applicable to sensors that are possibly cross-reactive and show non-linear responses. Specifically, we focus on the detection and design of MMSK schemes, which is a form of synthetic olfaction-inspired MC originally proposed in [4]. Natural olfaction relies on the detection of complex molecule mixtures, enabling plants and animals to perform robust communication, resource localization, and respond to environmental changes. While cross-reactivity has already been considered in [4], [5], these works do not consider practical non-linear macroscale sensors. On the other hand, non-linearities of MOS sensors were considered in [2] but no molecule mixtures. In the following, we introduce the considered system model in Sec. II, describe our proposed detector in Sec. III and a complementing mixture design algorithm in Sec. IV, evaluate the aforementioned detector and algorithm in Sec. V, and present some possible future research directions in Sec. VI.

## II. System Model

We consider an MMSK system, where a single transmitter (TX) transmits symbols $\mathbf{x}_i$ of a signal alphabet $\mathcal{X}_\mathrm{c}$ comprising $M$ symbols, where each entry in $\mathbf{x}_i$ corresponds to the released amount of one species. Each symbol $\mathbf{x}_i$ lies in some set of feasible concentrations $\mathcal{X} \subset \mathbb{R}^S_{\geq 0}$, where $\mathbb{R}_{\geq 0}$ and $S$ denote respectively the set of non-negative real numbers and the dimension of $\mathbf{x}_i$. The shape of $\mathcal{X}$ depends on the operating regime of the employed transmit and sensing units and optionally some molecule constraints. The concentration of molecules around the RX, denoted by $\mathbf{y}_i$, is given by $\mathbf{y}_i = \mathbf{h} \odot \mathbf{x}_i + \mathbf{n}(\mathbf{x}_i)$. Here, $\mathbf{h} \in \mathbb{R}^S_{\geq 0}$, $\mathbf{n}(\mathbf{x}_i)$, and $\odot$ denote respectively the known and fixed species-specific channel attenuation, a generally symbol-dependent noise vector, and the element-wise multiplication operator. In practice, the receiver (RX) cannot directly observe $\mathbf{y}_i$ but has to rely on the imperfect, non-linear output of sensors. We assume that the output $\mathbf{z}_i$ of the $n_\mathrm{SU}$ employed sensing units is described by a non-linear mapping $\mathbf{z}_i = \mathbf{f}(\mathbf{y}_i)$, so that the $k$-th entry in $\mathbf{z}_i$ corresponds to the output of the $k$-th sensor, which generally can depend on any entry in $\mathbf{y}_i$. While we assume negligible sensor noise here, our detector and mixture design algorithm can be readily extended to account for sensor noise.

## III. Symbol Detection for Non-Linear Receiver Arrays

Generally, it is tedious and computationally expensive to design and to employ optimal detectors for non-linear RXs, especially in cases without closed-form expression for $\mathbf{f}(\cdot)$, e.g., when $\mathbf{f}(\cdot)$ is obtained from data. Thus, we instead propose a low-complexity suboptimal detector based on the so-called unscented transform (UT) [6] requiring only knowledge of the first- and second-order moments of $\mathbf{y}_i$ and the ability to compute the output of $\mathbf{f}(\cdot)$ for all relevant inputs.

We employ the UT, as described shortly below, to estimate for each symbol $\mathbf{x}_i \in \mathcal{X}_\mathrm{c}$ the mean $\boldsymbol{\mu}_{\mathbf{z}_i}$ and covariance $\mathbf{C}_{\mathbf{z}_i}$ of the sensor output $\mathbf{z}_i$ as $\hat{\boldsymbol{\mu}}_{\mathbf{z}_i}$ and $\hat{\mathbf{C}}_{\mathbf{z}_i}$, respectively. These estimates are then used to parametrize some probability density function (pdf) $p_\mathrm{A}(\mathbf{z}_i|\hat{\boldsymbol{\mu}}_{\mathbf{z}_i}, \hat{\mathbf{C}}_{\mathbf{z}_i})$, which approximates the true likelihood $p(\mathbf{z}_i|\mathbf{x}_i)$. One candidate for this *approximate* pdf is the multivariate normal distribution as it is (i) readily parametrized by its mean and covariance, (ii) the most conservative choice in terms of Fisher information [7], and (iii) a common noise model for air-based MC [2]. Then, the transmitted symbol can be estimated as $\hat{\mathbf{x}} = \arg\max_{\mathbf{x}_i \in \mathcal{X}_\mathrm{c}} p_\mathrm{A}(\mathbf{z}_i|\hat{\boldsymbol{\mu}}_{\mathbf{z}_i}, \hat{\mathbf{C}}_{\mathbf{z}_i})$.

In the following, we briefly describe how the UT computes $\hat{\boldsymbol{\mu}}_{\mathbf{z}_i}$ and $\hat{\mathbf{C}}_{\mathbf{z}_i}$ given the mean $\boldsymbol{\mu}_{\mathbf{y}_i}$ and covariance $\mathbf{C}_{\mathbf{y}_i}$ of $\mathbf{y}_i$. The UT is based on $n_\sigma$ carefully selected points (*sigma points*) that have mean $\boldsymbol{\mu}_{\mathbf{y}_i}$ and covariance $\mathbf{C}_{\mathbf{y}_i}$. These points are fed through $\mathbf{f}(\cdot)$ and the transformed points are used to estimate the statistics of $\mathbf{z}_i$. We use the approach from [6] to choose $n_\sigma = 2S$ sigma points, which lie on the covariance ellipse around $\boldsymbol{\mu}_{\mathbf{y}_i}$. Specifically for the $i$-th symbol $\mathbf{x}_i$, the $j$-th and $(j+S)$-th sigma points are chosen as $\mathbf{s}_{i,j} = \boldsymbol{\mu}_{\mathbf{y}_i} + \sqrt{S\left[\mathbf{C}_{\mathbf{y}_i}\right]_j}$ and $\mathbf{s}_{i,j+S} = \boldsymbol{\mu}_{\mathbf{y}_i} - \sqrt{S\left[\mathbf{C}_{\mathbf{y}_i}\right]_j}$, respectively, where $\left[\mathbf{C}_{\mathbf{y}_i}\right]_j$ denotes the $j$-th column of $\mathbf{C}_{\mathbf{y}_i}$. Then, $\boldsymbol{\mu}_{\mathbf{z}_i}$ is estimated as $\hat{\boldsymbol{\mu}}_{\mathbf{z}_i} = \frac{1}{n_\sigma}\sum_{l=1}^{n_\sigma} \mathbf{f}(\mathbf{s}_{i,l})$ and $\mathbf{C}_{\mathbf{z}_i}$ is estimated as $\hat{\mathbf{C}}_{\mathbf{z}_i} = \frac{1}{n_\sigma}\sum_{l=1}^{n_\sigma} \left(\mathbf{f}(\mathbf{s}_{i,l}) - \hat{\boldsymbol{\mu}}_{\mathbf{z}_i}\right)\left(\mathbf{f}(\mathbf{s}_{i,l}) - \hat{\boldsymbol{\mu}}_{\mathbf{z}_i}\right)^\mathrm{T} + \mathbf{C}_0$, where $\mathbf{C}_0$ is a regularization parameter stabilizing the detector.

## IV. Molecule Mixture Design for Non-Linear Receiver Arrays

In the following, we propose a simple algorithm to design an MMSK alphabet $\mathcal{X}_\mathrm{c}$ for non-linear RXs which optimizes the separability of sensor outputs $\mathbf{z}_i$ as defined by the distance metric in [4, Eq. (7)]. This metric depends only on the mean and covariance of the sensor outputs $\mathbf{z}_i$.

We initialize the alphabet $\mathcal{X}_\mathrm{c}$ with a symbol $\mathbf{x}_1$ randomly drawn from $\mathcal{X}$. Then, we create a set of $n_\mathrm{cand}$ *candidate* mixtures

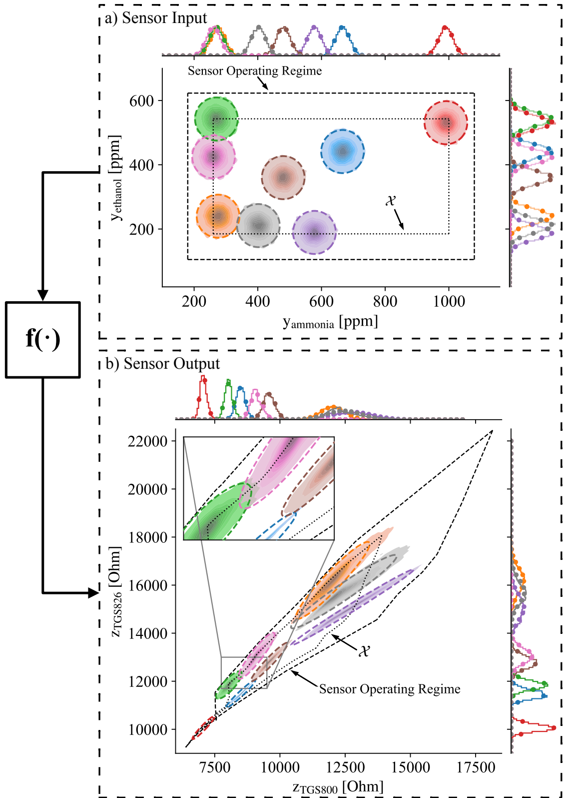


Fig. 1. **Exemplary Alphabet.** The empirical pdfs (darker colors indicate higher density) of the selected signal points are shown together with the $3\sigma$-confidence ellipses obtained from $\hat{\boldsymbol{\mu}}_{\mathbf{z}_i}$ and $\hat{\mathbf{C}}_{\mathbf{z}_i}$ (dashed ellipses). The marginal empirical distributions (solid), and the normal distributions obtained from $\hat{\boldsymbol{\mu}}_{\mathbf{z}_i}$ and $\hat{\mathbf{C}}_{\mathbf{z}_i}$ (dots) are shown on top and to the right in both the *sensor input domain* (top) and *sensor output domain* (bottom). We chose $\mathcal{X}$ so that also the noisy $\mathbf{y}_i$ lie in the sensor operating regime, i.e., the support of $\mathbf{f}(\cdot)$.

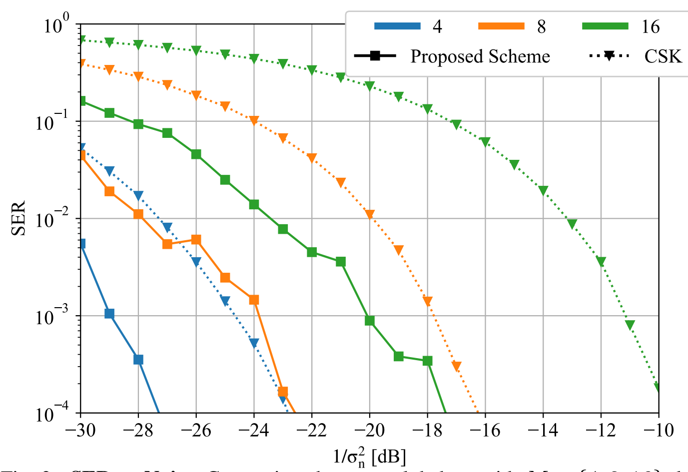


Fig. 2. **SER vs Noise.** Comparison between alphabets with $M \in \{4, 8, 16\}$ that are designed using our algorithm or the corresponding $M$-ary CSK-style scheme.

uniformly distributed in $\mathcal{X}$, compute the distance(s) between each candidate mixture and the point(s) already in $\mathcal{X}_c$, and add the one candidate, which has the largest minimum distance to any point already in $\mathcal{X}_c$. This process is iterated until $\mathcal{X}_c$ contains the desired number of symbols. Because the resulting alphabet is different for each execution of the afore-described algorithm, we run the algorithm $n_r$ times and keep only the $n_b$ alphabets with the largest minimum distance between all pairs of symbols.

## V. Evaluation

In this section, we present simulation results to demonstrate the performance of our proposed detector and mixture design algorithm. We use linearly interpolated measurements from [3], where the resistance of the TGS800 and TGS826 MOS sensors for different mixtures of ethanol and ammonia has been reported. We set $\mathbf{h}$ to an all-ones vector to focus on the reception process, assume zero-mean Gaussian independent and identically distributed noise with fixed variance $\sigma_n^2$, and choose $\mathbf{C}_0$ as a diagonal matrix with entries $10^3$.

Fig. 1 shows an exemplary alphabet with $M = 8$ for $\sigma_n^2 = 5 \cdot 10^2$ and $n_{\text{cand}} = 500$ in both the sensor input domain (top) and the sensor output domain (bottom). Clearly, the empirical pdfs of $\mathbf{y}_i$ match well the $3\sigma$ confidence ellipses obtained from $\hat{\boldsymbol{\mu}}_{\mathbf{z}_i}$ and $\hat{\mathbf{C}}_{\mathbf{z}_i}$ in both domains. Additionally, we show the empirical marginal distributions together with the approximating marginal pdfs parametrized by the corresponding entries in $\hat{\boldsymbol{\mu}}_{\mathbf{z}_i}$ and $\hat{\mathbf{C}}_{\mathbf{z}_i}$, demonstrating the accuracy of our approach. For a more quantitative understanding of the performance, Fig. 2 shows the symbol error rates (SERs) for $M \in \{4, 8, 16\}$, $n_r = 10$, $n_b = 5$, and $n_{\text{cand}} = 100$. We compare our approach to a CSK-style approach, where the symbols are equally spaced in $\mathcal{X}$ for ethanol[1] only while the ammonia concentration is fixed to $720\text{ppm}$. Clearly, our proposed approach, which exploits a two-dimensional design space, outperforms the CSK-style approach, which has only one design dimension, for all alphabet sizes by a significant margin, demonstrating the good performance of our detector and mixture design algorithm[2].

## VI. Conclusion and Future Work

In this work, we proposed an MMSK symbol detector and a molecule mixture design algorithm for non-linear RXs. So far, we evaluated our framework using sensor data from [3] for mixtures of ammonia and ethanol. In the future, we plan to extend our detector and algorithm to account for sensor noise. We also intend to perform experiments with molecules and sensors specifically chosen for MC applications. To achieve good detection performance, it is crucial to select molecules that exhibit high volatility, i.e., they can transition into the gas phase at room temperature, and the molecules should be robust enough to not react with other substances in the air.


## Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK 2950 – Project-ID 509922606.


## References


[1] S. Lotter *et al.*, "Experimental research in synthetic molecular communications – Part II," *IEEE Nanotechnol. Mag.*, vol. 17, pp. 54–65, June 2023.

[2] N.-R. Kim *et al.*, "A universal channel model for molecular communication systems with metal-oxide detectors," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1054–1059, 2015.

[3] S. Hirobayashi *et al.*, "Verification of a logarithmic model for estimation of gas concentrations in a mixture for a tin oxide gas sensor response," *Sens. Actuators B Chem.*, vol. 92, pp. 269–278, July 2003.

[4] V. Jamali *et al.*, "Olfaction-inspired MCs: Molecule mixture shift keying and cross-reactive receptor arrays," *IEEE Trans. Commun.*, vol. 71, pp. 1894–1911, Apr. 2023.

[5] M. O. Araz *et al.*, "Ratio shift keying modulation for time-varying molecular communication channels," *IEEE Trans. Nanobiosci.*, vol. 23, pp. 176–189, Jan. 2024.

[6] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, pp. 401–422, Mar. 2004.

[7] S. Park *et al.*, "Gaussian assumption: The least favorable but the most useful," *IEEE Signal Process. Mag.*, vol. 30, pp. 183–186, May 2013.


[1] We chose to vary the ethanol concentration since, compared to ammonia, it has a larger response range of the individual sensors.

[2] The slight performance deterioration at $1/\sigma_n^2 = -26\text{dB}$ for $M = 8$ is due to the random construction of $\mathcal{X}_c$, which is performed for each noise level separately.