Ultra-broadband infrared gas sensor for pollution detection: the TRIAGE project

To cite this article: Bruce Napier et al 2021 J. Phys. Photonics 3 031003

View the article online for updates and enhancements.
Ultra-broadband infrared gas sensor for pollution detection: the TRIAGE project

Bruce Napier1,*, Ole Bang2, Christos Markos3, Peter Moselund1, Laurent Huot1, Frans J M Harren1, Amir Khodabakhsh1, Hans Martin1, Flavia Ottonello Brianio4, Laurent Balet5, Steve Lecomte6, Christian R Petersen1, Niels Israelsen7, David Bastviken8, Magnus Gålfalk8, Łukasz Kubiszyn9 and Piotr Warzybok9

1 Vivid Components, Paderborn, Germany
2 DTU Fotonik, Technical University of Denmark, Lyngby, Denmark
3 NKT Photonics A/S, Birkerød, Denmark
4 Trace Gas Research Group, IMM, Radboud University, Nijmegen, The Netherlands
5 Senseair AB, Delsbo, Sweden
6 CSEM SA, Neuchâtel, Switzerland
7 NORBLIS IVS, Virum, Denmark
8 Department of Thematic Studies—Environmental Change, Linköping University, Linköping, Sweden
9 VIGO System S.A., Ozarow Mazowiecki, Poland

* Author to whom any correspondence should be addressed.
E-mail: bruce@vividcomponents.co.uk

Keywords: trace gas detection, supercontinuum source, spectroscopy, mid-infrared, deep learning algorithms, big data repositories

Abstract
Air pollution is one of the largest risk factors for disease or premature death globally, yet current portable monitoring technology cannot provide adequate protection at a local community level. Within the TRIAGE project, a smart, compact and cost-effective air quality sensor network will be developed for the hyperspectral detection of gases which are relevant for atmospheric pollution monitoring or dangerous for human health. The sensor is based on a mid-infrared supercontinuum source, providing ultra-bright emission across the 2–10 µm wavelength region. Within this spectral range, harmful gaseous species can be detected with high sensitivity and selectivity. The spectroscopic sensor, which includes a novel multi-pass cell and detector, enables a smart robust photonic sensing system for real-time detection. With built-in chemometric analysis and cloud connection, the sensor will feed advanced deep-learning algorithms for various analyses, ranging from long-term continental trends in air pollution to urgent local warnings and alerts. Community-based distributed pollution sensing tests will be verified on municipal building rooftops and local transport platforms.

1. Current challenges in air pollution detection

Air pollution, emitted by both anthropogenic and natural sources, constitutes a significant risk factor for a number of severe health conditions such as lung cancer and strokes. According to the World Health Organization, ambient air pollution (i.e. outdoor) was estimated to cause 4.2 M premature deaths worldwide and contributed to 7.6% of all deaths in 2016 [1]. The total annual economic cost of air pollution related health impacts is estimated to be >1.3 T€ [2]. These health concerns are driving increased national and international regulation on the monitoring and control of air pollution (figure 1). As a result, significant effort is devoted globally to improve air quality through, e.g. land-use planning strategies, waste management, replacement of fossil fuels by clean energy sources and lower levels of industrial and agricultural emissions. In order to be successful, these measures need to be accompanied by large scale air quality monitoring networks to ensure real-time citizen alerts on local pollution levels and compliance with air quality legislation. This is particularly challenging in dense urban areas with many local emission sources, leading to complex air mixtures which makes simple risk analysis difficult.
1.1. Limitations of existing systems
Currently, several methods are applied for air quality monitoring, among them mass spectrometry, various types of low cost sensors (sometimes combined to constitute electronic noses) and optical detection. Most of the analytical systems based on mass spectrometry are highly sensitive and selective but, while some progress has been made in developing portable systems, due to their vacuum technology they suffer from low mobility and high cost. Sensors and electronic noses are low cost but lack accuracy and are difficult to maintain due to the wide variety of sensor types needed for different compounds, each with particular maintenance requirements [3, 4]. In contrast, optical methods can in principle offer accurate simultaneous identification and concentration measurement of many specific species in a complex mixture, while at the same time achieving high sensitivity in real-time. Multiple systems based on coherent, single frequency light sources operate in the infrared (IR) wavelength range, such as diode, quantum cascade and interband cascade lasers, as well as nonlinear approaches using optical parametric oscillation or difference frequency generation [5–7]. However, they do not operate in a hyperspectral mode, covering only a limited range of wavelengths. Frequency combs typically cover a wide wavelength range and enable fast, highly sensitive and selective detection of molecules [8–10]. Such high performance tools generally remain confined to academic research laboratories due to their operational complexity and prohibitively high cost.

To enable smart, distributed sensor networks and community-based real-time monitoring and warning systems for air pollution requires new solutions which are more precise, robust and reliable than present instrumentation, and which can simultaneously quantify multiple air pollutants, whose chemical fingerprints are spread across a wide spectral range.

1.2. The TRIAGE project
The EU Horizon 2020 project TRIAGE (uITRa-broadband InFrAred Gas sEnsor) aims to deliver novel instrumentation that can provide the basis for large sensor networks for the comprehensive monitoring of 17+ air pollutants, greenhouse gases and toxic molecules [11], where toxicity level is defined via the Immediately Dangerous to Life and Health exposure scale [12]. TRIAGE will thereby contribute to a safer environment by providing detailed air quality data near community areas, industrial infrastructure, highways, harbours, rural areas and in the case of catastrophic events e.g. chemical accidents. The portable sensor will provide long-term pervasive monitoring data on predefined individual molecular species, and also for the identification of a priori unexpected substances, generating a complete picture of air quality composition.

TRIAGE will be able to provide data in the same way as other ground-based optical remote sensing methods, such as light detection and ranging (LIDAR) or differential optical absorption spectroscopy, but with much higher selectivity and sensitivity for molecular gases. The sensor provides a cost-effective means for rapid, real-time air quality sampling with high mobility, which is compatible for use within large sensing networks. It aims to provide crucial data ranging from real-time citizen alerts to long-term trends in air pollution monitoring.

1.3. TRIAGE system concept
The general working principle is shown in figure 2: a mid-IR supercontinuum (SC) source covering the wavelength region between 2 and 10 µm will propagate (figure 3) through an environmental air sample with an optical path length of many metres in either a compact multi-pass cell or through free space e.g. between
buildings using a retroreflector. In both cases, a compact Fourier transform IR (FTIR) spectrometer will be used in combination with an IR detector for hyperspectral analysis. The data will be analysed using sophisticated machine learning algorithms and neural networks (i.e. deep learning), based on a vast database of training sets in order to provide accurate information on the species and pollution levels (figure 4). The concentration results and original data will be available in a cloud-based environmental database.
2. TRIAGE system hardware

2.1. Mid-IR supercontinuum source
The mid-IR wavelength range is often referred to as the 'fingerprint' region, due to the presence of strong and distinct fundamental rotational–vibrational transitions within molecular gases. Diode-pumped, optical fibre-based SC sources can cover a large part of this wavelength region with high power and brightness [13]. As SC sources emit all wavelengths simultaneously, the challenges of wavelength tuning, timing and stabilisation are avoided. Moreover, despite covering a wide wavelength range, the spectral brightness is remarkable: the shoe-box sized mid-IR SC source outperforms thermal sources, currently in use with commercial FTIR spectrometers, by 4–5 orders of magnitude and even synchrotron sources by one order of magnitude [14, 15]. The SC source provides a near diffraction-limited, spatially coherent beam, which is ideal for use over long path lengths. As such they are one of the most suitable mid-IR sources for hyperspectral chemical sensing applications.

The broad spectrum is realised by the gradual broadening of the input pump spectrum through a series of optical fibres with increasing mid-IR transparency. The efficiency of this process, also known as cascaded SC generation, requires that the optical fibres are matched along the cascade to allow for efficient coupling and strong confinement [16]. To this end, a SC source will be developed to increase output power using tailored chalcogenide fibres with strong confinement and anti-reflection nano-imprinted glass fibre surfaces [17].

Most relevant sources are based on 1.55 µm pump lasers that are widely available, due to their applications in telecom and LIDAR. However, these lasers are very energy inefficient and the process used to shift the power to 2 µm and beyond is complex, and can generate significant noise. In recent years, thulium-doped fibre laser technology has advanced rapidly, enabling much more efficient power generation at 2 µm. This technology will be applied to develop a much more efficient low noise pump laser for the SC source and a total output power of >100 mW is expected in the wavelength range 2–10 µm.

2.2. Mid- and long-wavelength IR detectors
TRIAGE will develop dedicated detection modules which integrate IR detectors, Peltier coolers and front-end electronics in sealed miniaturised packages to be used as a part of the FTIR system. The active element of the single pixel IR detection module will be an InAs/InAsSb superlattice (SL)-based heterostructure grown on a buffered IR-transparent substrate. The heterostructure will consist of multiple layers with proper doping profile and SL period composition. The signal-to-noise performance will be improved by the use of enhanced absorption measures and monolithically integrated micro-lenses. The front-end electronics will ensure proper operating conditions for the detectors and a low noise linear gain of the electric signal.

2.3. Spectrometer
For the mid-IR spectral analysis, TRIAGE will optimise an FTIR spectrometer design [18], which will be adapted to be compact, fast scanning, transportable and able to accommodate the 2–10 µm SC source. As all components are COTS (commercial off-the-shelf), the FTIR is low cost. The FTIR spectrometer will have a spectral resolution of 1 GHz (∼0.033 cm⁻¹) and a single shot measurement time of a few seconds over the complete wavelength range, thanks to the improved SC source and detectors.

2.4. Multi-pass cell
Commercial multi-pass cells will not be used, because they are too expensive, delicate, labour intensive and complicated to interface with the other system blocks. In order to realise a high quality opto-mechanical system that can also be produced efficiently in large quantities, a customised, rugged and lightweight absorption cell will be designed and developed using precision moulding technology [19]. It will embed electronics, pressure and temperature sensors as well as mirror heaters to reduce condensation, thus reducing assembly cost and paving the way for cost-efficient production.

3. Spectral analysis and machine learning
The advantage and strength of gas sensing in the mid-IR wavelength region lies in its extreme sensitivity and specificity for detecting traces of molecular gases in seconds. The detection sensitivity strongly depends on the overlap between the wavelength coverage of the SC source and the strongest absorption lines of the molecular gases of interest. The specificity depends on the spectral overlap between the absorption of the studied gas and other gases present.

In order to monitor continuously the concentration of the targeted volatiles, real-time data analysis algorithms are needed. Since the calculations involved are highly demanding in terms of processing power, TRIAGE will use a cloud-based approach. The quantification of multiple gas species with overlapping
absorptions is much more challenging than single species detection. A broadband absorbance spectrum can be covered by taking a large number of spectral elements within the wavelength range. This method is particularly beneficial to decompose partially overlapping absorption features. In practice, a priori knowledge of the number of target gas species and their absorbance patterns is required for the algorithms to work. As such, a reference database will be taken for a large number of gaseous volatile compounds. The absorption spectra of a vast number of gases can be found in the PNNL \cite{20} and the HITRAN \cite{21} databases.

3.1. Initial data processing
Within TRIAGE, the initial data processing will apply partial least square fitting methods, in combination with factor analysis on a created training set that contains the natural variation to be expected in the samples to be analysed. Via this approach, to some extent, non-linearities are accounted for as an extra factor in the basis set (baseline drift, mirror degradation, complexes of individual species, reaction products etc) \cite{22}.

3.2. Deep learning methods
Several machine learning approaches will be investigated to estimate gas concentrations from absorption spectra. Traditional methods, such as support vector machines and Gaussian processes, will be considered as they have been shown to perform well in many regression tasks. In a second phase, convolutional neural networks (CNNs) will be evaluated for the task of estimating gas concentrations. CNNs can learn to extract high level features from the raw input data. In contrast, traditional approaches usually require handcrafted features to achieve good performance.

To create a training set, a large dataset of real atmospheric spectra, collected over a range of path lengths, temperatures, pressures, spectrometer variations, and background variations of water vapour and carbon dioxide will be assembled. Known reference spectra of target compounds will be added digitally to extend the calibration training set to species not usually present in a clean atmosphere.

4. Planned experimental demonstration

4.1. Building on the FLAIR project
In order to speed up the development of the algorithms, real datasets will be required in addition to database-generated spectra. Since the TRIAGE prototype will not be ready during the early phases of the project, the 2–10 $\mu m$ prototype SC-based system built during a previous H2020 project (FLAIR: Flying ultrA-broadband single shot InfraRed sensor) will be employed, as shown in figure 5.

4.2. TRIAGE field trials
The evaluation of system performance for monitoring, early warning, and prediction will be performed in coordination with ongoing air quality monitoring in cities and regions in several participant countries. Some examples include monitoring of emissions from waste incineration, biogas production and handling, and city air quality in e.g. Linköping, Stockholm and Neuchâtel. Field trials of city air quality will involve local transport platforms such as buses, making pollution monitoring possible across a whole city at high resolution in both space and time.

TRIAGE will contribute to the evaluation of local and regional efforts to reach national environmental goals regarding air quality, including information sharing between these nodes, and coordination of alerts, e.g. through cooperation with the Swedish Environmental Protection Agency, the Swedish Institute for Hydrology and Meteorology, Stockholm Air and Noise Analysis and the Service de l’énergie et de l’environement from the canton Neuchâtel.
4.3. Conclusion
The ground-breaking potential of the TRIAGE sensor system is expected to have a considerable impact on air quality monitoring, providing fundamentally new capacity. TRIAGE aims to overcome the hurdle which to date has hindered the widespread application of the mid-IR wavelength region, by applying new technology, both hardware and software, with the potential to reduce costs by an order of magnitude and at the same time provide performance beyond state-of-the-art. The portable system will offer improvement in selectivity and sensitivity for a wide range of pollutant species, with real-time analysis; data previously obtainable only with high-end laboratory systems.

Data availability statement
No new data were created or analysed in this study.

Acknowledgments
TRIAGE (Grant No. 101015825) and FLAIR (Grant No. 732968) have received funding from Horizon 2020, the European Union’s Framework Programme for Research and Innovation.

ORCID iDs
Bruce Napier https://orcid.org/0000-0001-7223-7478
David Bastviken https://orcid.org/0000-0003-0038-2152

References
[1] WHO Fact Sheet 2018 Ambient (outdoor) air pollution (available at: www.who.int/news-room/fact-sheets)
[2] WHO Regional Office for Europe, OECD 2015 Economic cost of the health impact of air pollution in Europe: clean air, health and wealth