Automatic processing of multi-modal chemical tomography datasets

A. D. Parsons[a], *S. Price*[a] , *A. M. Beale*[bc], *M. Basham*[a], *N. Wadeson*[a], *J. F. W. Mosselmans*[a], *P. D. Quinn*[a]

[a]*Diamond Light Source, Didcot ,OX11 0DE*
[b]UK Catalysis Hub, Research Complex at Harwell, Didcot OX11 0FA
[c]University College London, London, WC1H 0AJ

Biological and material science problems are often investigated by a single microscopy/ spectroscopy technique. However, such investigations using a single technique often yield deeper questions that require further investigation. The continuing development of chemical tomography techniques, yielding spatially resolved information on e.g. elemental and phase distribution, is able to provide a more detailed picture of the nature of a material than the corresponding bulk measurements.

One example of a system that benefits from multi technique analysis is the determination of the active state of a catalyst. The physicochemical state of a material is a key factor in determining catalytic activity and selectivity, however rarely are such materials structurally or compositionally homogeneous. The use of X-ray fluorescence computed tomography (XRF-CT) is able to identify the location of the active state, but not its crystal structure or chemical state. For this, the collection of X-ray diffraction computed tomography (XRD-CT) or X-ray absorption near edge structure computed tomography (XANES-CT) is required[1]. These datasets are collected in parallel, not only reducing the duration (and hence required dose) of the experiment, but also guaranteeing the sample was in the same state for each measurement, something not possible is an dynamic industrial catalyst. This ultimately provides data for correlative image processing.

Processing time is a major consideration when designing an experiment to collect high quality data from such a multi-modal experiment, since the number of processing hours soon exceeds that which is viable for a user to spend processing them. This also detracts from the possibility of real-time analysis as the experiment progresses, so users are often left until they leave the beamline to assess how successful the experiment was. Ideally, the user would be able to view the reconstructed data as the experiment progresses, enabling them to guide the experiment on the fly.

Big data pipelines are already having a significant impact on the data analysis world outside of science, but are limited to low dimensionality data-sets. Savu[2] is a python based data processing pipeline developed at DLS that is targeted specifically at the higher dimensionality of scientific data-sets. Scientists can write simple plugins to the pipeline that can then be applied in a modular manner to the data to reduce it to its final state dramatically improving the processing time of the data.

We present the results of benchmarking the performance improvement versus a standard, user-based processing chain for this example data-set, demonstrating a dramatic increase in processing time and provide scope for future improvements as we push computationally intensive data processing towards real-time feedback.

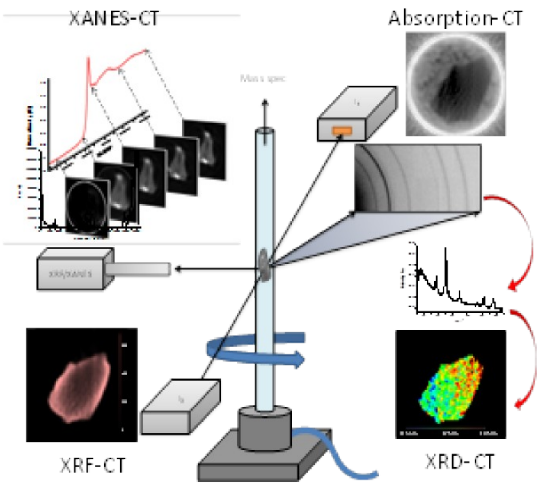

Figure 1: The experimental set-up for a multimodal absorption, XRF, XANES and XRD contrast CT, measurement of an industrial catalyst.

References

[1] Price et al. Phys. Chem. Chem. Phys., 2015, 17, 521--529

[2] Atwood et al. Phil. Trans. R. Soc. A 373: 20140398