IEF Final Report

MARIE CURIE ACTIONS

Intra-European Fellowships (IEF)

PIEF-GA-2009-253063-CVM-EM-PALM

"Computing the structure and dynamics of protein assemblies in living cells by coupling sub-diffraction fluorescence microscopy with single-particle reconstruction: application to viral capsids and nuclear pores"

Reporting Period: 1.5.2010 - 31.11.2011

Researcher: Claas von Middendorff, PhD

Project Coordinator: Christophe Zimmer, PhD

Project Coordinator Organization Name :

Institut Pasteur, Paris, France

Address of the Researcher:
Claas von Middendorff, PhD
Computational Imaging and Modeling Group
(G5 Groupe Imagerie et Modélisation)
Institut Pasteur
25-28 rue du Dr Roux
75724 Paris
Tel: 0033 (0)1 40 61 33 77
E-mail: cvmidden@pasteur.fr

Address of the Scientist in Charge:
Christophe Zimmer, PhD
Computational Imaging and Modeling Group
(G5 Groupe Imagerie et Modélisation)
Institut Pasteur
25-28 rue du Dr Roux
75724 Paris
Tel: 0033 (0)1 40 61 38 91
E-mail: czimmer@pasteur.fr

URL:http://www.pasteur.fr/ip/easysite/go/03b-00001l-023/unit-s-et-groupes/g5-imagerie-et-mod-lisation

Final Publishable Summary Report

Our objective is to develop computational procedures to enhance image sharpness in super-resolution fluorescence microscopy by adequate modeling of the spatial and spatio-temporal dynamics of the marker molecule.

Context. Image sharpness is a central issue when medical or biochemical questions are studied by fluorescence microscopy. In the last ten years, modern fluorescence microscopes have been invented that feature a resolution that is not anymore limited by the diffraction of light. Hereby, the image sharpness could be increased by roughly one order of magnitude allowing researchers to investigate many biological objects of high importance that were hidden to the observer before. Specifically, methods like Stimulated Emission Depletion Microscopy (STED) or Stochastic Reconstruction Microscopy (STORM) arrive routinely at resolving objects of the size of $\approx 20nm$ with visible light and far-field optical equipment. This has opened up a completely new field of research at the crossing of physics, engineering, biochemistry and informatics modeling: super-resolution fluorescence microscopy.

A particulary important subclass of the mentioned super-resolution methods relies on switching randomly the fluorescence capability of the marker molecules, classical optical camera detection of the activated molecule and subsequent single-molecule localization. With random fluorescence switching, only a sparse aleatoric subset of otherwise very densely packed molecules is observed. For an isolated molecule the camera image is typically an Airy disc of $\approx 200nm$ diameter. The aforementioned aleatoric subset of activated molecules is likely to be sufficiently sparse in order to contain few overlapping single-molecule images. From these latter the true molecule position can be estimated to sub-diffraction accuracy depending only on the photons-to-noise ratio of the image, i.e. the strength of the light stream emitted by the isolated molecule compared to the background. These high-resolution position estimates can be used to form a super-resolved image just by tabulating the number of position estimates falling into a certain spatial volume. Doing this, the approach reaches a resolution of a few tens of nm.

Objectives and Results. Our main scientific question is : **Can the resolution**

be pushed even more, to the single-digit nanometer range? In the project, we propose a step towards achieving this by exploiting the fact that every marker molecule is usually detected and localized in many fluorescence switching cycles. If the information of the different detections is combined, a composite position estimate should be obtainable that is much more precise than the constituent single-cycle detections. This implies that resolution in localization-based super-resolution microscopy depends not only on the photon number of the individual molecule detection event, but also on its **detection multiplicity**. This effect is

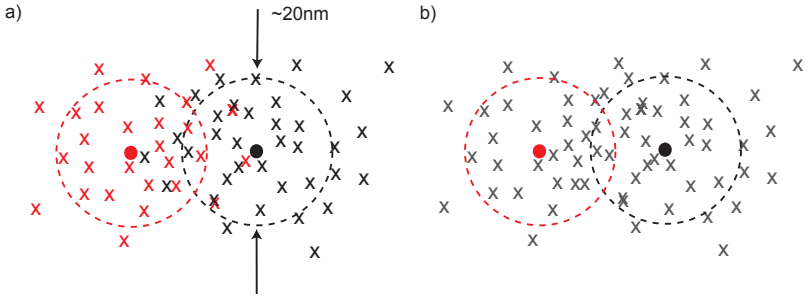

Figure 1: A schematic model for the spatial detection behavior for two fluorescent marker molecules (black, red). For every molecule there are up to thousands of position estimates (crosses) that fluctuate with an error of roughly $20nm$ (dashed circles) around the true positions (big dots). Averaging these estimates would lead to single-digit nanometer accuracy for the position of the molecule. However, in reality the assignment of position estimates to its originator molecules is not fully clear (a) but rather ambiguous (b). This ambiguity can by partly removed by exploiting prior knowledge about the positioning error probability distribution. During the Marie Curie project a mathematical method to do this, the REEP method, has been developed thereby allowing to improve image sharpness considerably.

illustrated in Figure 1. In order to combine the multiple position estimates originating from the same molecule, the researcher has developed **algorithms for a computational procedure** called **Resolution Enhancement by Estimate Pooling (REEP)** that allows to cluster position estimates in space and time based on the statistical signature of the individually detected molecules. To this aim, methods from technical computing, data mining, stochastic modeling and unsupervised learning have been combined. Particularly, we have found **mathematical formulae that describe the spatial and spatio-temporal clustering** of molecular position detections. In our image model, the **spatial distribution of position estimates** of a collection of K molecules at original positions μ_k is

represented by a Gaussian mixture model,

$$p(\mathbf{r}|\mu_k) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{r}|\mu_\mathbf{k}, \Sigma_k),$$

that can 'learn' its clustering distribution via the expectation-maximization algorithm. In turn, to describe the **temporal distribution of position estimates** we have calculated the probability that a molecule with mean on and off times $\bar{\tau}_{on}, \bar{\tau}_{off}$ and the mean number of fluorescence switching cycles $\bar{M}$ is detected in an image frame at time t as

$$p_d(t|\bar{M}, \bar{\tau}_{on}, \bar{\tau}_{off}) = \sum_M p(M|\bar{M}) \int_{\tau>t}^{\infty} \{Erlang(\tau, \bar{\tau}_{on}, M) * Erlang(\tau, \bar{\tau}_{off}, M-1)\}\, d\tau.$$

Furthermore, we created a **virtual microscope simulation environment**. This latter predicts a considerable sharpness increase on a system of outmost biological importance, the nuclear pore, see Figure 2. The nuclear pore system as an assembly object of a few ordered molecules has led us to formulate **special resolution measures** that quantify the fidelity with which the pooling procedure assigns position estimates to its originator molecule. The notions of uniqueness and miscellaneousness complement the usual notion of resolution as localization accuracy of single molecules. The whole groundwork is inspired by **single-particle reconstruction** - a method from electron microscopy where large numbers of copies of identical particles are superposed in order to enhance the image contrast.

In the Marie-Curie project we have (i) investigated how multi-particle superposition can create detection multiplicity, (ii) explored how to mathematically merge individual detections in order to obtain a composite estimate, (iii) devised the computer algorithm package REEP to perform this task, (iv) tested the performance of the algorithms as far as the resolution issue is concerned, especially to which extent the details of nuclear pores could be resolved and (v) experimentally demonstrated the REEP procedure on two-molecule DNA origami nanoscopic ruler data.

Scientific Training. During the project the researcher has been trained in modern computational data mining procedures, particularly supervised learning, regression, shrinkage methods, classification, kernel smoothing, model selection, model averaging, ensemble learning and neural networks.

Impact. High-resolution microscopy is a mature technique by now and images are produced routinely in many laboratories around the world. In this setting of availability of vast amounts of imaging data it is expected that the concept of the Marie-Curie project - image sharpness enhancement in single-marker switching microscopy by running statistical learning methods on existing data sets - contributes considerably to widen the scope of nano-sized biological objects

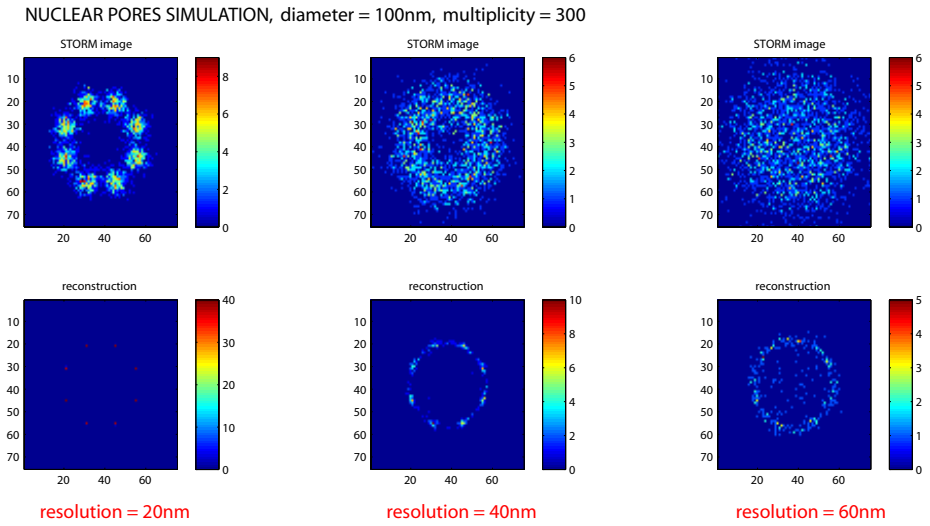

Figure 2: Monte-Carlo simulation for the imaging process of a nuclear pore complex with the conventional high-resolution STORM image (top row) post-processed using the REEP algorithm (bottom row) for a nuclear pore of diameter $d = 100nm$, i.e. nucleoporine distance $\approx 40nm$. The number of detections per molecule - the multiplicity - is $\bar{M} = 300$. The underlying STORM resolution is varied as $r = 20, 40, 60nm$ (from left to right). Note that for the STORM resolution $r = 40nm$ our algorithm allows to resolve the eight molecules of the nuclear pore complex whereas they stay hidden in the STORM microscopy image.

that could be studied. By making the technique available in an environment of biomedical basic research like the Pasteur Institute, the new method is likely to be useful for medical researchers producing thereby a direct benefit for society.

The researcher has organized a lectures series on high-resolution microscopy with members of his former laboratory that contributed to a knowledge transfer between the Max-Planck-Institute for Biophysical Chemistry, Germany and the Pasteur Institute, France. Thereby, and by attending the conference 'Focus on Microscopy', April 2011, efforts were made to involve other actors, to spread awareness and to disseminate the project results.

Keywords: *high-resolution fluorescence microscopy, single-marker switching, sub-diffraction imaging, scientific computation, mixture models, stochastic imaging models, estimate pooling, mathematical resolution enhancement, super-resolution image processing, single-particle reconstruction, nuclear pore, DNA origami*