



Lukas Schmid

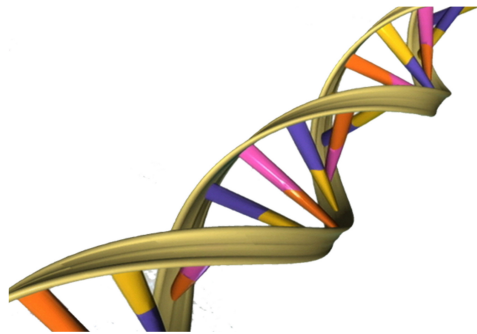


Roman Koller

Graduate Candidates	Lukas Schmid, Roman Koller
Examiner	Prof. Dr. Guido Schuster
Co-Examiner	..
Subject Area	Sensor, Actuator and Communication Systems
Project Partner	Confidential

# All your Base for Next Generation DNA Sequencing

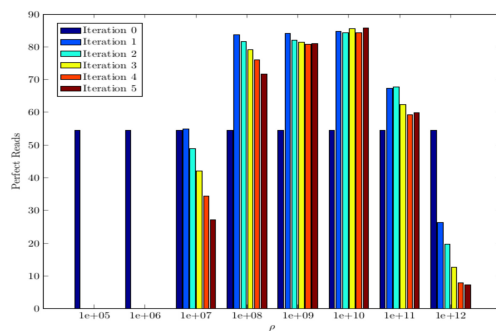
## Term Thesis



DNA double helix

**Introduction:** Massingham and Goldman published an article about an iterative algorithm called «All your Base» (AYB). It applies a promising statistical approach for base calling. This is an important process in next generation sequencing, i.e. it is the actual classification of the nucleobases.

**Approach/Technologies:** This thesis gives detailed explanations and derivations of the algorithm including its reimplementation in Matlab. A mathematical equation is set up which models the most important effects, phasing, fading and crosstalk, into the input intensities. Using the Tikhonov Regularization helps to estimate parameters between its initial value and a least-square solution. After removing these unwanted interferences the actual base calling procedure is applied. Two versions are implemented, the simple base call which finds the maximum value and a Viterbi base calling which is much more complicated. It consists of computing the inverse covariance matrix using a maximum-likelihood approach with a model based on a Gaussian distribution and the Polak-Ribiere and Newton-Raphson methods. Finally, a Hidden Markov Model is established and the Viterbi algorithm can be applied to find the most likely base sequences.



This figure shows an improvement of the outputs over five iterations with rho=10<sup>10</sup>.

**Result:** Extensive experiments are able to reproduce the results of the paper and more importantly, to evaluate the algorithm on raw image data by a new product. Since the regularization method requires initialization the algorithm relies very much on these inputs. Furthermore, tests have shown that estimating the interaction parameters heavily depends on the regularization parameter. An additional feature has been developed within this thesis that lowers the sensitivity of this parameter significantly. Moreover, it could have been surprisingly demonstrated that Viterbi does not improve the quality notably when prototype data is used.