Scientific writing

Ivan Viola

& collective

UNIVERSITETET I BERGEN

Forms of Information Dissemination

- Scientific Paper
  - Journal publication
  - Conference publication (Proceedings)
- Poster
- Thesis (Master, PhD)
- Tutorial/Course Notes
- STAR-Report
- Invited talk
- Popular reports in news and broadcasting
- X-Mas Card?

Important Journals and Conferences

- Journals
  - IEEE Trans. on Visualization and Computer Graphics (TVCG)
  - IEEE Computer Graphics and Applications (CG&A)
  - Transactions on Graphics (TOG)
  - Computer Graphics Forum (CGF)
- Conferences and Symposia
  - IEEE Visualization
  - Eurographics/IEEE VGTC Symposium on Visualization (EuroVis)
  - ACM SIGGRAPH
  - Eurographics

Forms of Information Dissemination

- Scientific Paper
  - Journal publication
  - Conference publication (Proceedings)

- Poster
- Thesis (Master, PhD)
- Tutorial/Course Notes
- STAR-Report
- Invited talk
- Popular reports in news and broadcasting
- X-Mas Card?

Scientific Publication

- Mostly used form of dissemination of novel ideas
- Target audience: specific scientific community
- Ensuring the quality: peer-reviewing, international program committee (IPC), program chairs
- Must read as a story
  - **Problem** (no overview on fish stocks, uncontrolled fishing)
  - **Goal** (sustainable fishery)
  - **Approach** (sonar scanning of sea regions and interactive visual analysis of data)
  - **Outcome** (precise quota estimation)
  - **Costs** (research, surveillance devices, manpower, ships)

Paper Quality Evaluation

- Scientific value
  - Does it address an important problem?
  - Level of novelty (significant step or incremental?)
  - Technical soundness
  - In scope of the journal/conference?
- Presentation quality
  - Clarity of presentation
  - Problem description
  - References the most important work in the field
  - Reproducability

Scientific Publication

- Often describes difficult and complex approach
- Reader should be considered!

  If the reader is to grasp what the writer means, the writer must understand what the reader needs.

  [G. Gopen and J. Swan 1990]

1. t(time)=15', T(temperature)=32º; t=0', T=25º; t=6', T=29º; t=3', T=27º; t=12', T=32º; t=9'; T=31º

2.

time (min)	temperature(ºC)
0	25
3	27
6	29
9	31
12	32
15	32

3.

temperature(ºC)	time (min)
25	0
27	3
29	6
31	9
32	12
32	15

Style Guidelines

- Always keep the **red thread**!
- Be consistent among terms (set-up, setup, set up)
- Reasoning first, approach details afterwards
- Avoid using contractions (e.g., it's, can't, ...)
- No paragraph should have only one sentence
- Use a spellchecker
- Sentence should not be too long (max 2 lines)
- Use simple formulations and short sentences, if a sentence can be broken into two, do it.
- Try to avoid abbreviations (except: i.e., e.g.), if more convenient, introduce them first
- Avoid colloquiality

Style Guidelines (cont.)

- Serif font is designed for easier reading the line
- Use sans-serif in titles but not in the plain text
- Include methodological illustrations, rendered images
- Equations variables have to be described in the text
- Equations, Figures, Tables all has to be referenced and discussed in the text labelled and having caption
- Captions - little redundancy is welcome
- Very important to explain first the high-level information and then to dig into the details
- Use consitency among terms
- Prefer positive formulations over negative ones

Paper Structure

- Title
- Authors
- Abstract
- Keywords, Index Terms

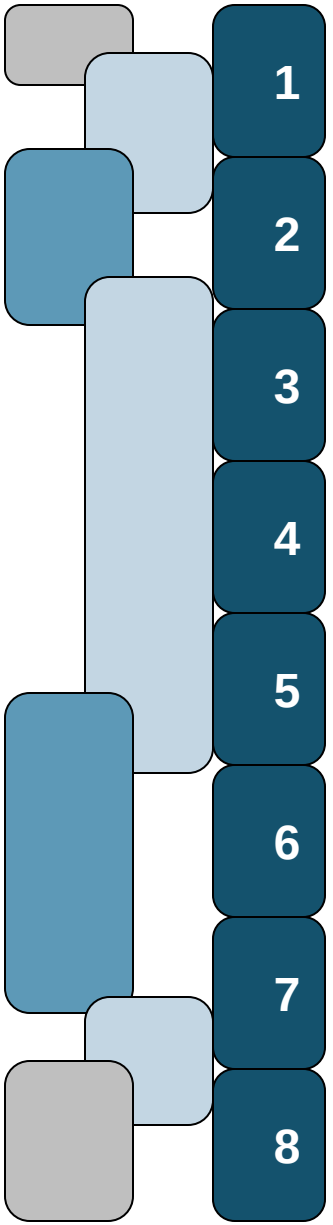
Head

- Introduction
- Related Work
- Basic Idea
- Approach in Detail
- Experiments, Validation,
- Results, Performance, Discussion
- Outlook
- Conclusions

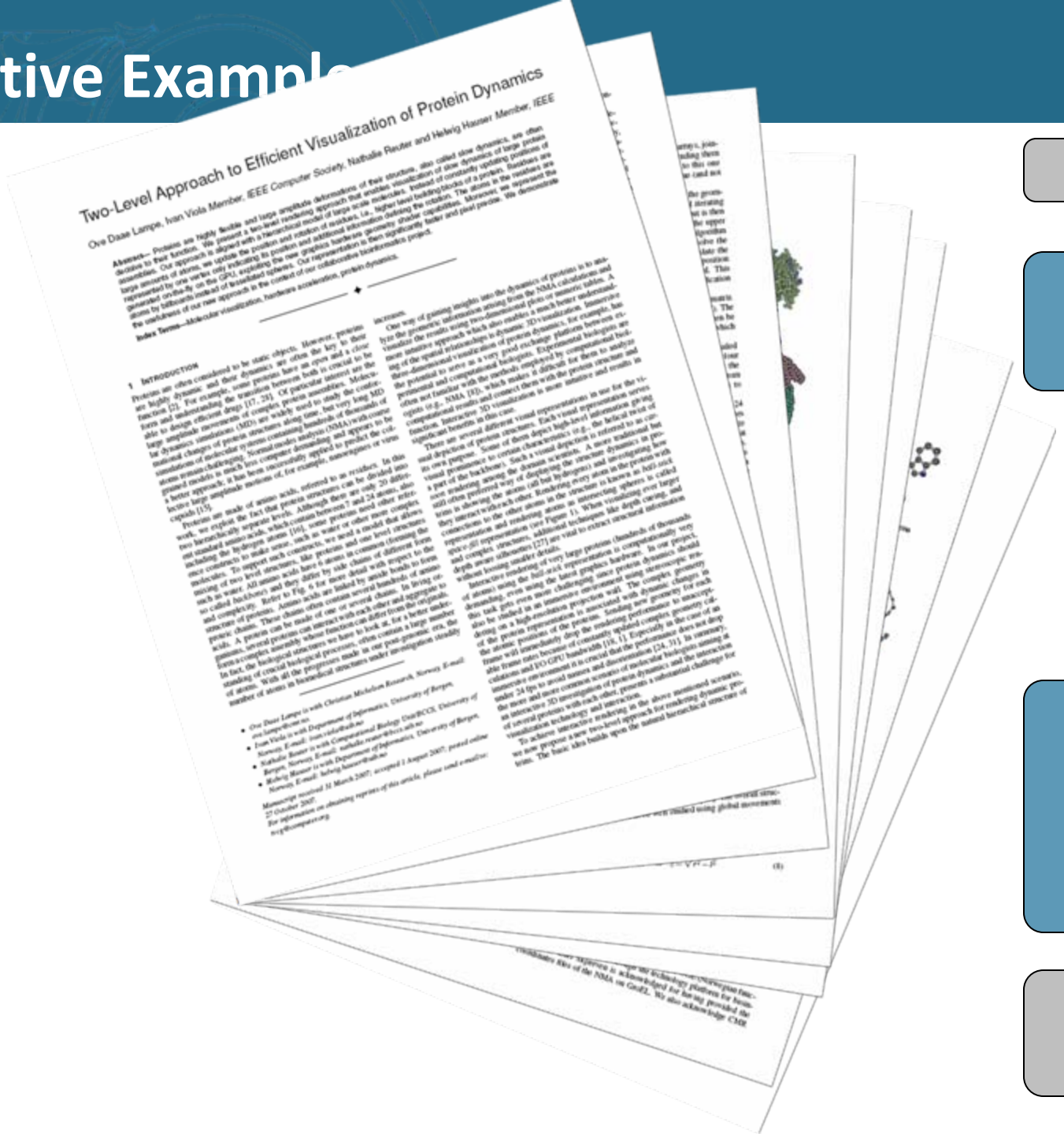
Body

- References, Appendix

Tail



Illustrative Example



Two-Level Approach to Efficient Visualization of Protein Dynamics

Ove Daae Lampe, Ivan Viola Member, IEEE Computer Society, Nathalie Reuter and Helwig Hauser Member, IEEE

Abstract—Proteins are highly flexible and large amplitude deformations of their structure, also called slow dynamics, are often decisive to their function. We present a two-level rendering approach that enables visualization of slow dynamics of large proteins consisting of several thousand secondary structure elements, i.e., higher-level building blocks of a protein. Rendering is performed by one vertex only exploiting the position and additional information during the rotation. The atoms in the meshes are generated on-the-fly on the GPU, exploiting the new graphics hardware geometry shader capabilities. Moreover, we represent the atoms by billboards instead of tessellated spheres. Our representation is then significantly faster and pixel precise. We demonstrate the usefulness of our new approach in the context of our collaborative bioinformatics project.

Index Terms—Molecular visualization, hardware acceleration, protein dynamics.

1 INTRODUCTION

Proteins are often considered to be static objects. However, proteins are highly dynamic and their dynamics are often the key to their function [2]. For example, some proteins have an open and a close form and understanding the transition between both is crucial to be able to design efficient drugs [17, 25]. Of particular interest are the large amplitude movements of complex protein assemblies. Molecular dynamics simulations (MD) are widely used to study the conformational changes of protein structures along time, but very long MD simulations of molecular systems containing hundreds of thousands of atoms remain challenging. Normal mode analysis (NMA) with coarse grained models is much less computer demanding and appears to be a better approach; it has been successfully applied to predict the collective large amplitude motions of, for example, nanoengines or virus capsids [15].

Proteins are made of amino acids, referred to as residues. In this work, we exploit the fact that protein structures can be divided into two hierarchically separate levels. Although there are only 20 different standard amino acids, which contain between 7 and 24 atoms, also nine constructs to make sense, such as water or other more complex molecules. To support such constructs, we need a model that allows mixing of two level structures, like proteins and ligands, but also such as water. All amino acids have 4 atoms in common (forming the so called backbone) and they differ by side chains of different form and complexity. Refer to Fig. 6 for more detail with respect to the structure of proteins. Amino acids are linked by amide bonds to form chains. These chains often contain several hundreds of amino acids. A protein can be made of one or several chains. In living organisms, several proteins can interact with each other and aggregate to form a complex assembly where function can differ from the originals. In fact, the biological processes, often contain a large number of atoms. With all the progresses made in our post-genomic era, the understanding of crucial biological processes under investigation steadily increases.

One way of gaining insights into the dynamics of proteins is to analyze the geometric information arising from the NMA calculations and visualize the results using two-dimensional plots of numeric tables. A more intuitive approach which also enables a much better understanding of the spatial relationships is dynamic 3D visualization. Immersive three-dimensional visualization of protein dynamics, for example, has the potential to serve as a very good exchange platform between experimental and computational biologists. Experimental biologists are often not familiar with the methods employed by computational biologists (e.g., NMA [8]), which makes it difficult for them to analyze computational results and connect them with the protein structure and results significant benefits in this case.

There are several different visual representations in use for the visual depiction of protein structures. Each visual representation serves its own purpose. Some of them depict high-level information giving visual prominence to certain characteristics (e.g., the helical twist of a part of the backbone). Such a visual depiction is referred to as cartoon rendering among the domain scientists. A more traditional but still often preferred way of displaying the structure dynamics in 3D could be drawing the atoms (all but hydrogen) and investigating how they interact with each other. Rendering every atom in the structure is called CPK representation, see Figure 1). When visualizing ever larger connections to the other atoms in the structure is known as ball-and-stick representation and rendering atoms as interacting spheres is called space-fill representation (see Figure 1). When visualizing ever larger and complex structures, additional techniques like depth cueing, and depth aware silhouettes [27] are vital to extract structural information without loosing smaller details.

Interactive rendering of very large protein (hundreds of thousands of atoms) using the full-scale representation is computationally very demanding, even using the latest graphics hardware. In our project, this task gets even more challenging since proteins should also be rendered in an immersive environment with stereoscopic rendering on a high-resolution projection wall. The complex geometry of the protein representation is associated with dynamic changes in the static positions of the proteins. Sending new geometry for each frame will immediately drop the rendering performance to unacceptable frame rates because of constantly updated complex geometry for calculations and I/O GPU bandwidth [18, 1]. Especially in the case of an immersive environment it is crucial that the performance does not drop under 24 fps to avoid nausea and disorientation [24, 31]. In summary, the more and more common scenario of molecular biologists aiming at an interactive 3D investigation of protein dynamics and the interaction of several proteins with each other, presents a substantial challenge for visualization technology and interaction.

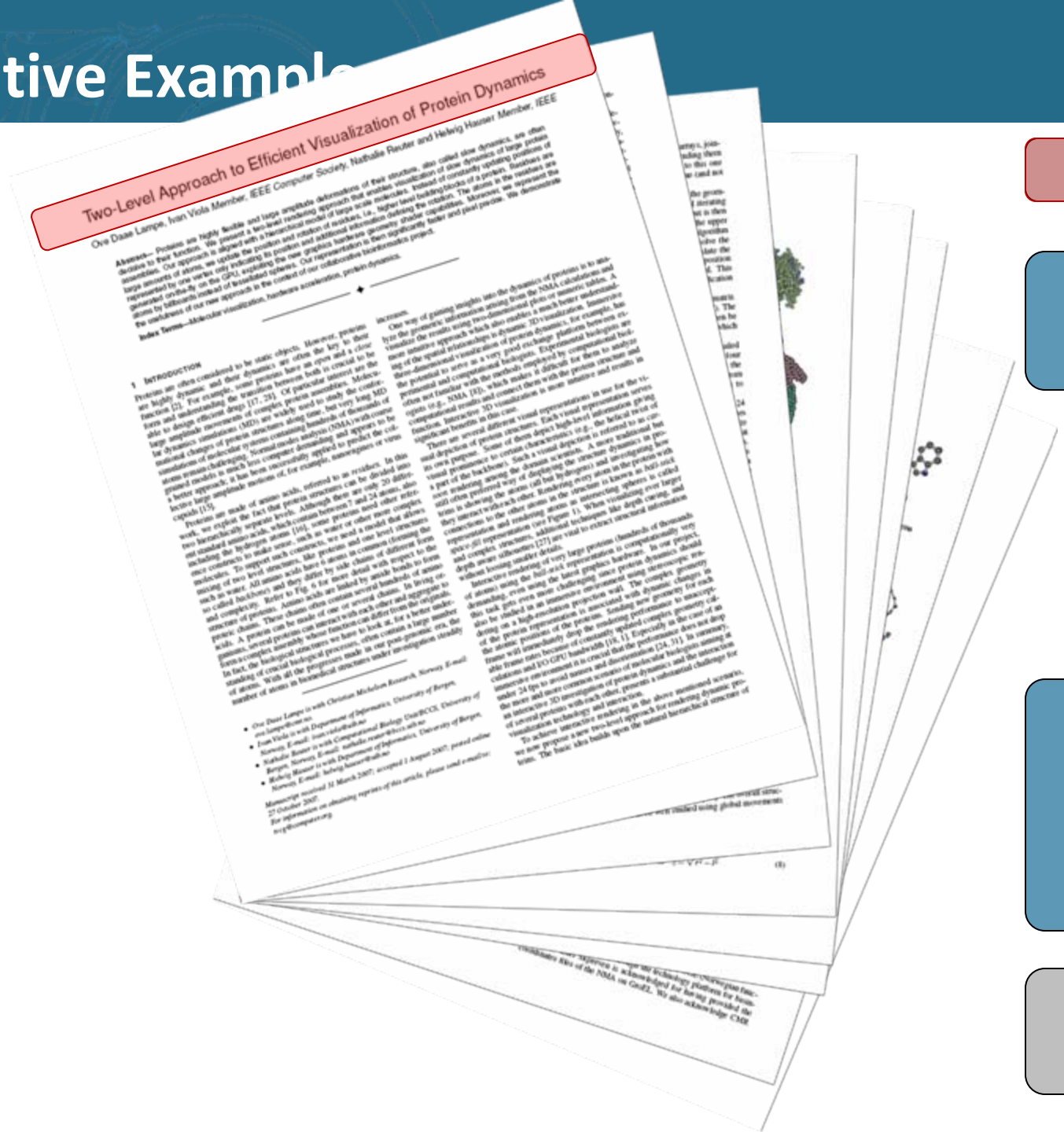
To achieve interactive rendering in the above mentioned scenario, we now propose a new two-level approach for rendering dynamic proteins. The basic idea builds upon the natural hierarchical structure of

- Ove Daae Lampe is with Christian Michelsen Research, Norway. E-mail: ove.lampe@cmr.no.
- Ivan Viola is with Department of Informatics, University of Bergen, Norway. E-mail: ivan.viola@ii.uib.no.
- Nathalie Reuter is with CoreModule, Norway. E-mail: nathalie.reuter@core.no.
- Helwig Hauser is with Department of Informatics, University of Bergen, Norway. E-mail: helwig.hauser@uib.no.

Manuscript received 31 March 2007; accepted 1 August 2007; posted online 22 October 2007.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

...Norway's Joint-Industry Project "Visualization of Complex Flow Structures" (VIZIOCOM). The authors would like to thank Mario Hlawitschka for the fruitful discussions and the anonymous reviewers for their comments. The molecular dynamics simulation of the NMA on Grid'5. We also acknowledge CSIC

Illustrative Example



1
2
3
4
5
6
7
8

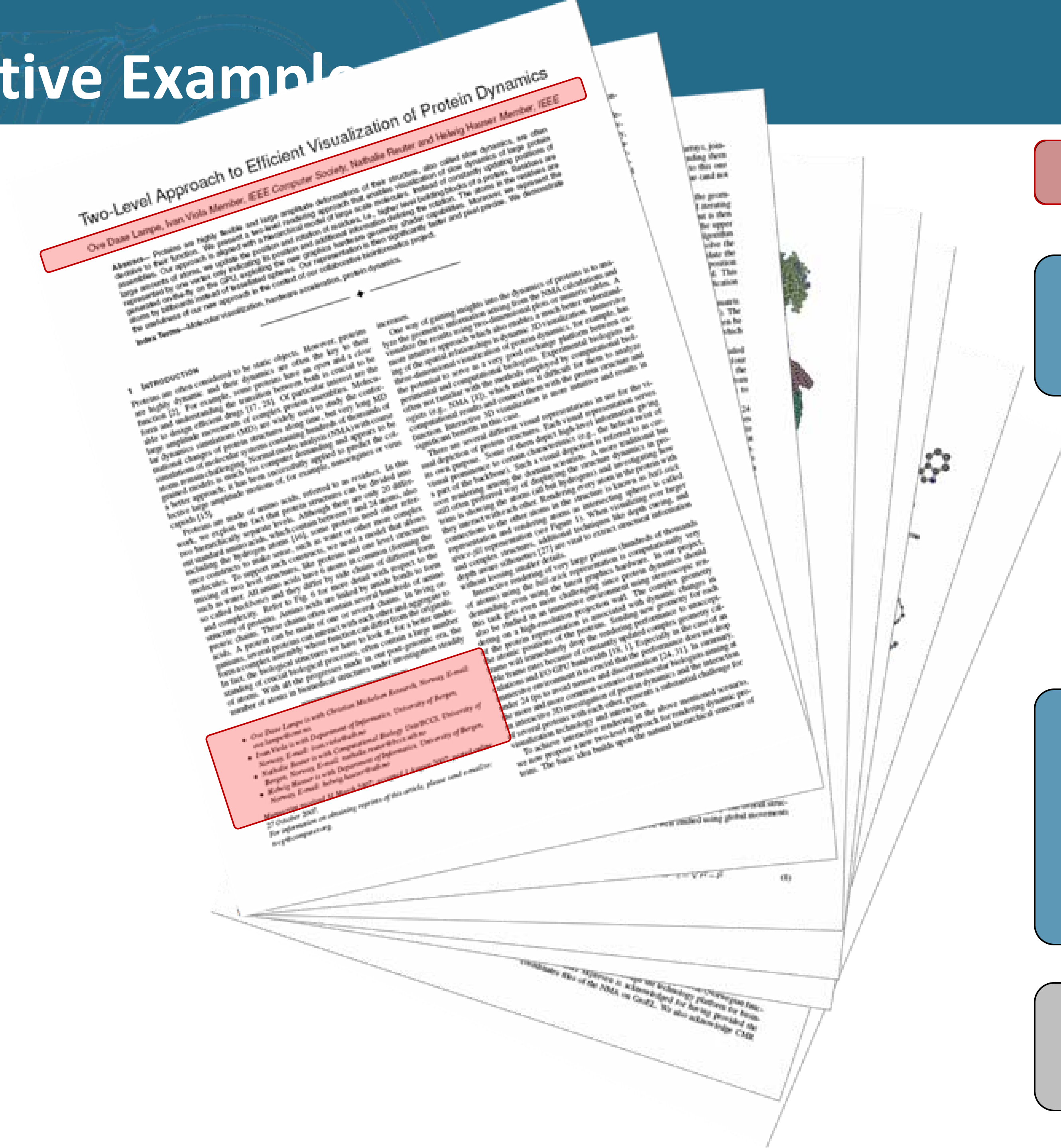
Title

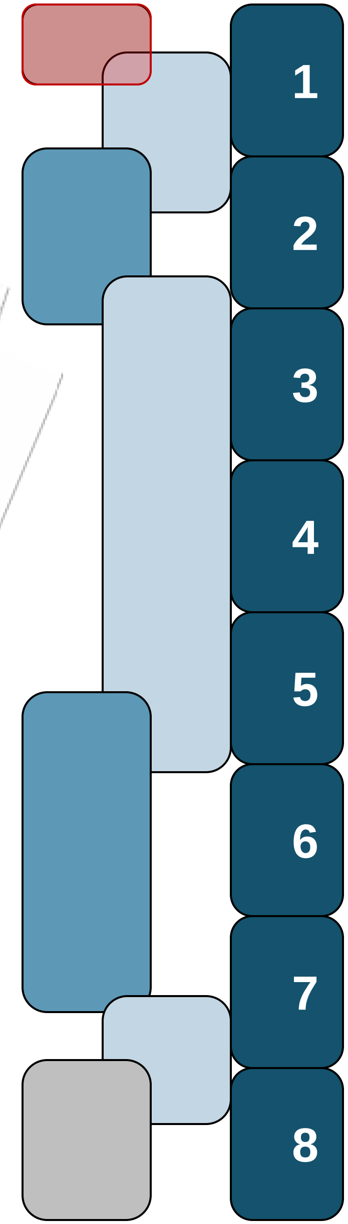
- **Thousands of people read the title without reading rest!**
- Quoting Master: ”Kurz, prägnant, arogant!“ (*short, concise, arogant*)
- Two (extreme) examples from IEEE Vis conference
  - ***Superellipsoid-based, Real Symmetric Traceless Tensor Graphs Motivated by Nematic Liquid Crystal Alignment Visualization***
    - Describes exactly what it is about but disctracts a reader right at the beginning
  - ***Caricaturistic Visualization***
    - Too general title, hard to find out what the paper specifically address
- Correct title design lies somewhere in between

uhhh brrrrr, wwwwwwhat?

hmm, sounds interesting...

Illustrative Example

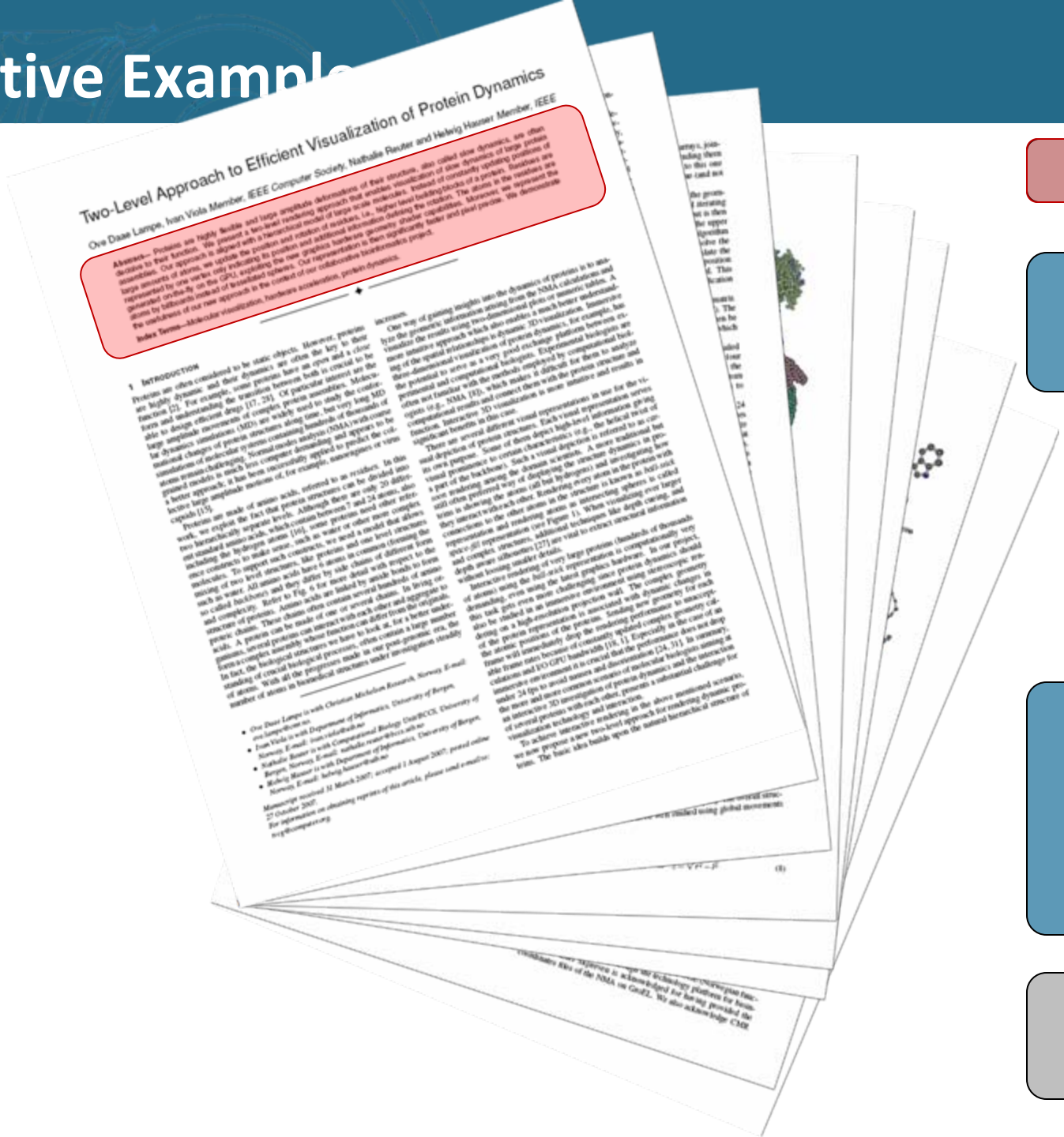




Authors and Affiliation

- Often several authors contribute to one publication
- First author: main responsible, main research contributor, work coordinator, trigger
  - Often PhD student, ...
- Co-authors → *et al.* (et alii, and others)
  - Project manager (PostDoc?)
  - Students
  - Supervisor (usually the last author)
- Affiliation – research institution
  - Several affiliations per person are allowed
  - Authors connected to affiliations through markups

Illustrative Example



Two-Level Approach to Efficient Visualization of Protein Dynamics

Ove Daae Lampe, Ivan Viola Member, IEEE Computer Society, Nathalie Reuter and Helwig Hauser Member, IEEE

Abstract— Proteins are highly flexible and large amplitude deformations of their structure, also called slow dynamics, are often decisive to their function. We present a two-level rendering approach that enables visualization of slow dynamics of large protein assemblies. Our approach is aligned with a hierarchical model of large scale molecules. Instead of constantly updating positions of large amounts of atoms, we update the position and rotation of nuclears, i.e., higher level building blocks of a protein. Transitions of nuclears are represented on-the-fly on the GPU, exploiting the rule and additional information defining the rotation. The atoms in the nearbies are generated on-the-fly on the GPU, exploiting the rule and additional information defining the rotation. The atoms in the nearbies are atoms by billboards instead of tessellated spheres. Our representation is then significantly faster and pixel precise. We demonstrate the usefulness of our new approach in the context of our collaborative bioinformatics project.

Index Terms—Molecular visualization, hardware acceleration, protein dynamics.

1 INTRODUCTION

Proteins are often considered to be static objects. However, proteins are highly dynamic and their dynamics are often the key to their function [2]. For example, some proteins have an open and a close form and understanding the transition between both is crucial to be able to design efficient drugs [17, 25]. Of particular interest are the large amplitude movements of complex protein assemblies. Molecular dynamics simulations (MD) are widely used to study the conformational changes of protein structures along time, but very long MD simulations of molecular systems containing hundreds of thousands of atoms remain challenging. Normal mode analysis (NMA) with coarse grained models is much less computer demanding and appears to be a better approach; it has been successfully applied to predict the collective large amplitude motions of, for example, nanomotors or virus capsids [15].

Proteins are made of amino acids, referred to as residues. In this work, we exploit the fact that protein structures can be divided into two hierarchically separate levels. Although there are only 20 different standard amino acids, which contain between 7 and 24 atoms, also more complex residues, such as water or other non-standard ones, including the hydrogen atoms [16], are used in some protein structures. To support such constructs, we need a model that allows grouping of two level structures, like proteins and ligands, into molecules. To support such constructs, we need a model that allows grouping of two level structures, like proteins and ligands, into molecules. Two atoms are bonded if they are linked by amide bonds to form such as water. All atoms are bonded if they are linked by amide bonds to form so called backbones and they differ by side chains of different form and complexity. Refer to Fig. 6 for more detail with respect to the structure of proteins. Amino acids are linked by amide bonds to form proteic chains. These chains often contain several hundreds of amino acids. A protein can be made of one or several chains. In living organisms, several proteins can interact with each other and aggregate to form a complex assembly where function can differ from the originals. In fact, the biological processes, often contain a large number of atoms. With all the progresses made in our post-genomic era, the understanding of crucial biological processes, often contain a large number of atoms. With all the progresses made in our post-genomic era, the number of atoms in hierarchical structures under investigation steadily increases.

One way of gaining insights into the dynamics of proteins is to analyze the geometric information arising from the NMA calculations and visualize the results using two-dimensional plots or numeric tables. A more intuitive approach which also enables a much better understanding of the spatial relationships in dynamic 3D visualization. Immersive three-dimensional visualization of protein dynamics, for example, has the potential to serve as a very good exchange platform between experimental and computational biologists. Experimental biologists are often not familiar with the methods employed by computational biologists (e.g., NMA [8]), which makes it difficult for them to analyze computational results and connect them with the protein structure and significant benefits in this case.

There are several different visual representations in use for the visual depiction of protein structures. Each visual representation serves its own purpose. Some of them depict high-level information giving visual prominence to certain characteristics (e.g., the helical twist of a part of the backbone). Such a visual depiction is referred to as cartoon rendering among the domain scientists. A more traditional but still often preferred way of displaying the structure dynamics in 3D could be drawing the atoms (all but hydrogen) and investigating how they interact with each other. Rendering every atom in the structure is known as ball-and-stick representation (see Figure 1). When visualizing even larger connections to the other atoms in the structure is known as ball-and-stick representation and rendering atoms as interacting spheres is called space-fill representation (see Figure 1). When visualizing even larger and complex structures, additional techniques like depth cueing, and depth aware silhouettes [27] are vital to extract structural information without loosing smaller details.

Interactive rendering of very large protein (hundreds of thousands of atoms) using the full-scale representation is computationally very demanding, even using the latest graphics hardware. In our project, this task gets even more challenging since the rendering should also be ported as an immersive environment with stereoscopic rendering on a high-resolution projection wall. The complex geometry of the protein representation is associated with dynamic changes in the static positions of the proteins. Sending new geometry for each frame with tens/hundreds of constantly updated complex is amongst calculations and I/O GPU bandwidth [18, 1]. Especially in the case of an immersive environment it is crucial that the performance does not drop under 24 fps to avoid nausea and disorientation [24, 31]. In summary, the more and more common scenario of molecular biologists sitting in an interactive 3D immersive cubic presents a substantial challenge for visualization technology and interaction.

To achieve interactive rendering in the above mentioned scenario, we now propose a new two-level approach for rendering dynamic proteins. The basic idea builds upon the natural hierarchical structure of

- Ove Daae Lampe is with Christian Michelsen Research, Norway. E-mail: ove.lampe@cmr.no.
- Ivan Viola is with Department of Informatics, University of Bergen, Norway. E-mail: ivan.viola@ii.uib.no.
- Nathalie Reuter is with ComBioMed, Computational Biology Unit/BCCS, University of Bergen, Norway. E-mail: nathalie.reuter@biocs.uib.no.
- Helwig Hauser is with Department of Informatics, University of Bergen, Norway. E-mail: helwig.hauser@uib.no.

Manuscript received 31 March 2007; accepted 1 August 2007; posted online 22 October 2007.

For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

This work is supported by the Norwegian Research Council through the VISIMOL project. We acknowledge the Bergen Center for Computational Science for having provided the computational resources of the NMA on Grid2. We also acknowledge CMR

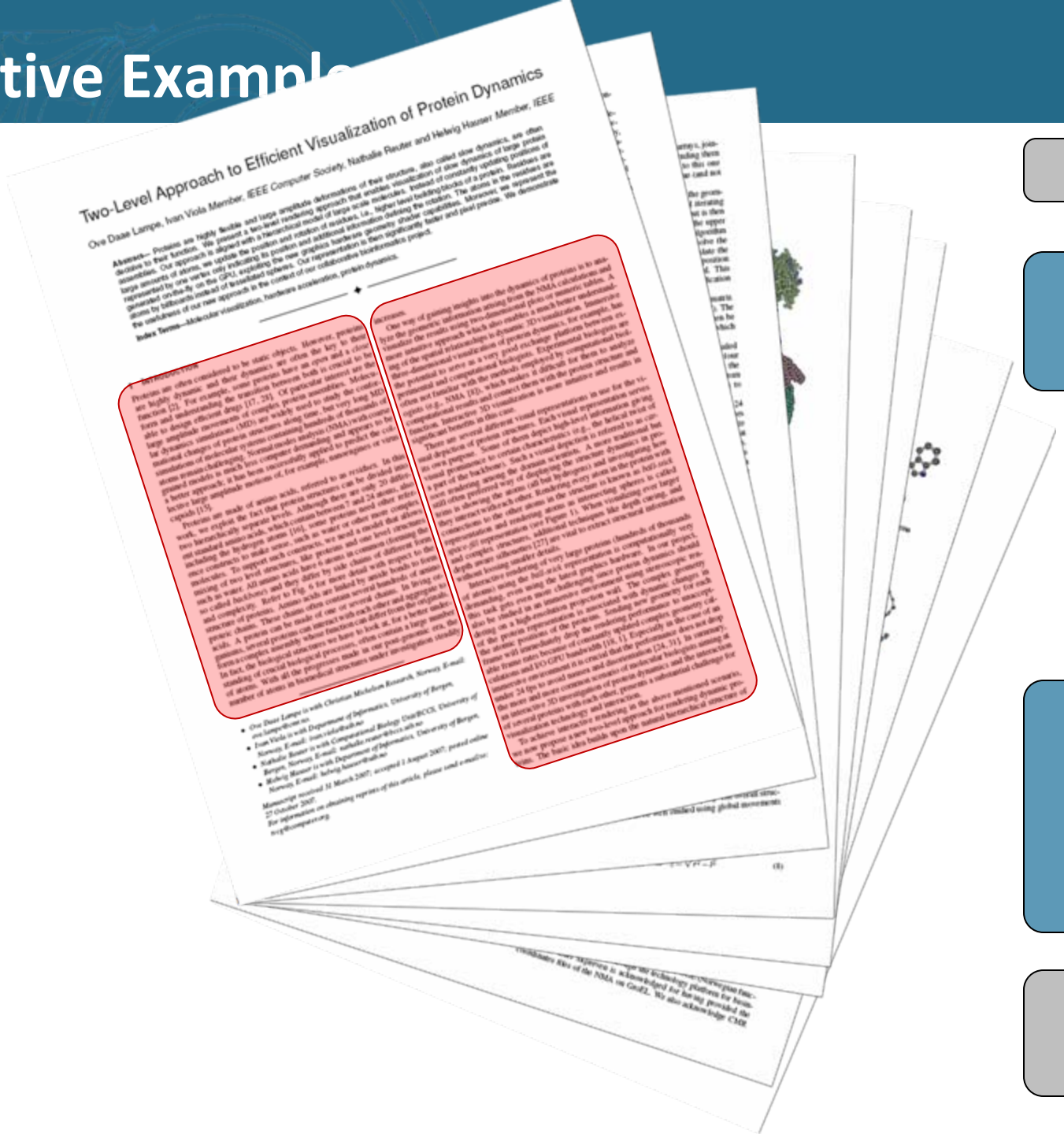
1
2
3
4
5
6
7
8

Abstract and Keywords

- Abstract includes
  - Questions you investigated
  - One sentence motivation for the work
  - Short description of the method
  - Advantages over existing methods
  - A brief summary of conclusions
- Ca 200 words not longer
- Motivate reader to read further
- Keywords
  - To enable easy indexing
  - Explicit indexing through classification system

http://www.acm.org/class/1998/

Illustrative Example

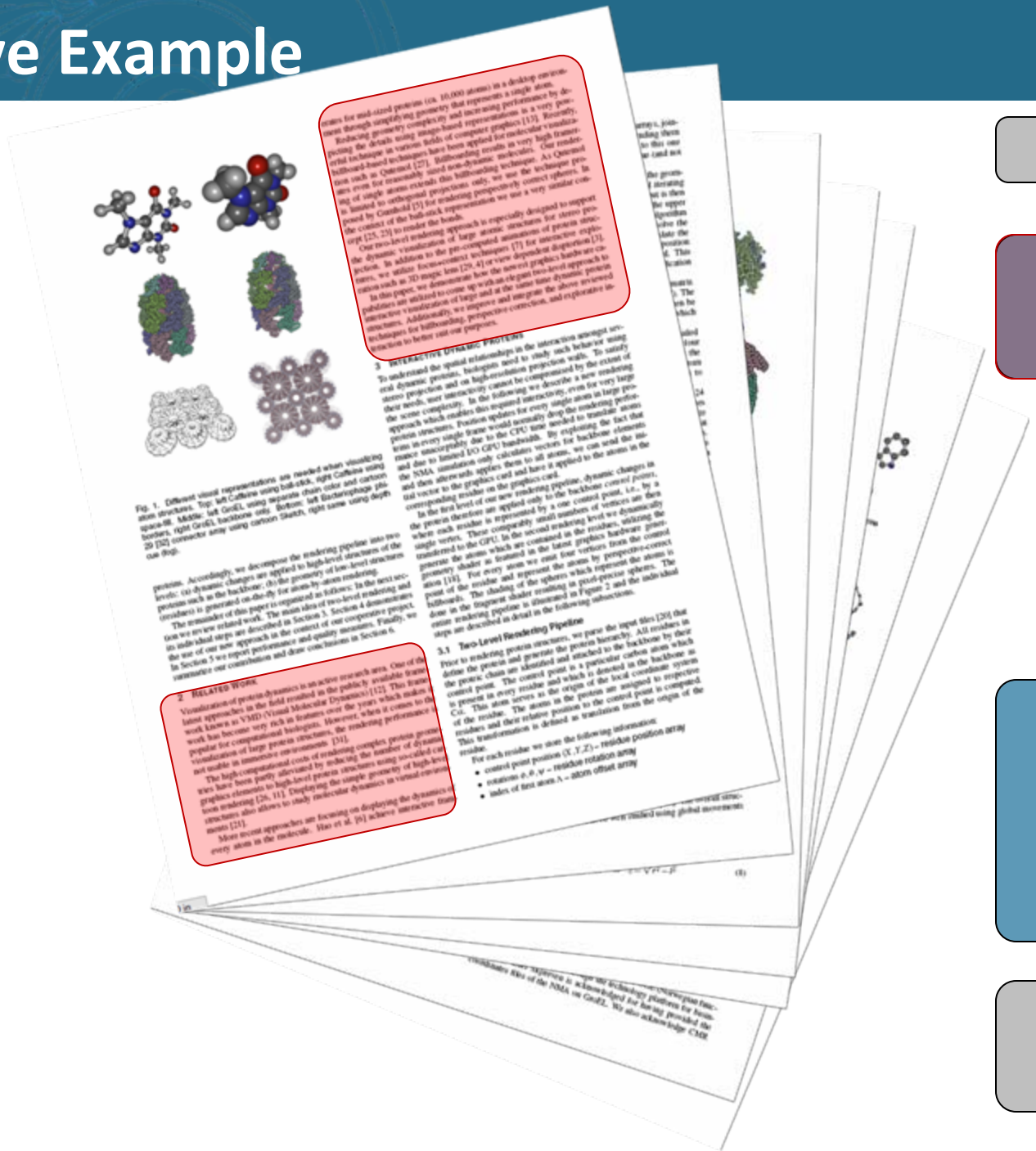


1
2
3
4
5
6
7
8

Introduction

- Background knowledge
  in visualization often application domain specifics
- Problem description
- Stating objectives
- Main contribution of your approach
- An nice resulting rendered image in the front is always good as an appetizer
  (relevant mostly for Vis and CG)

Illustrative Example

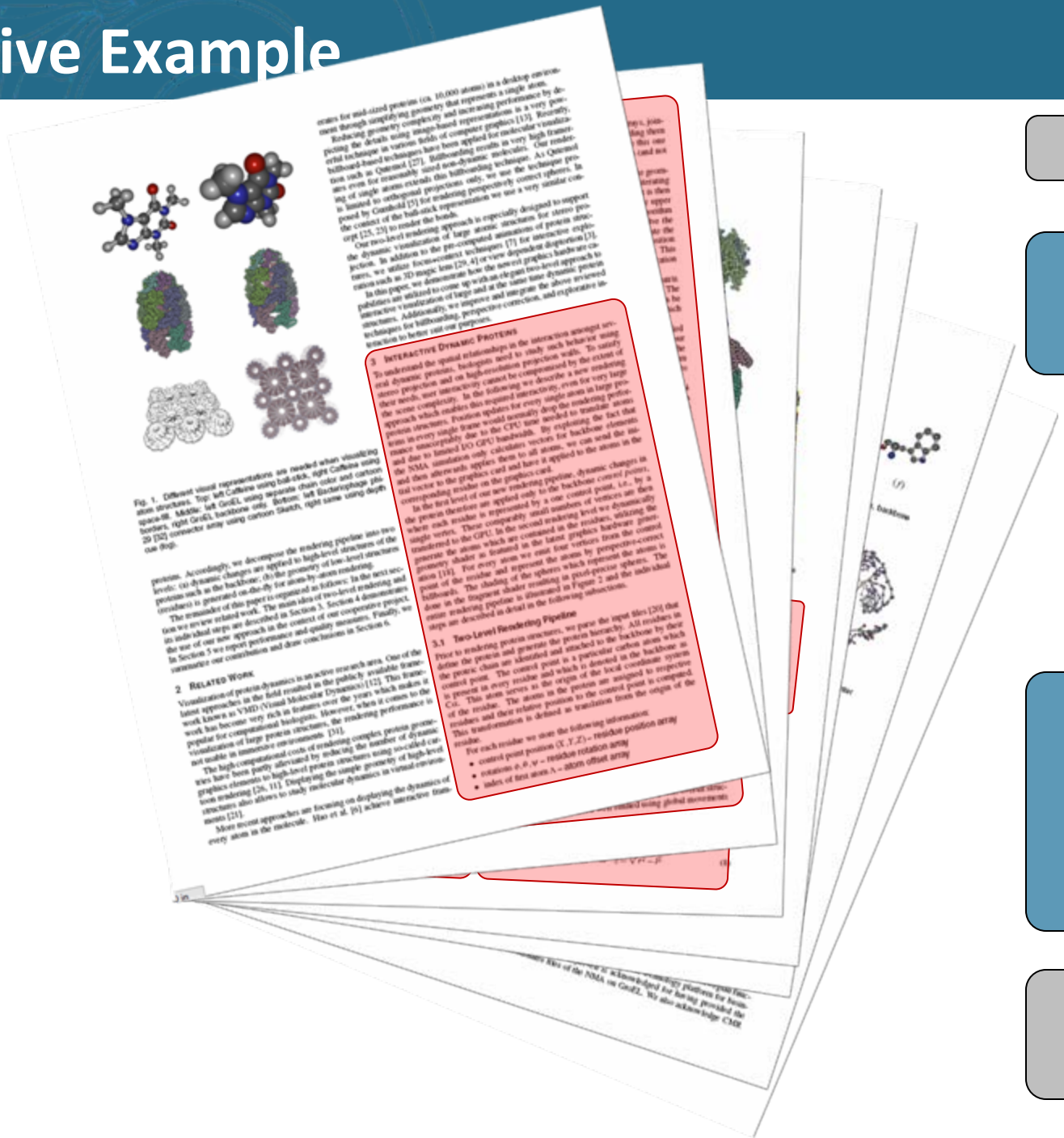


1 2 3 4 5 6 7 8

Related Work

- Put the work into context of other approaches
- Clearly state what has been known until now
- Cite every relevant work
- Provide brief comparison to your approach
- Citations writing style:
  - Two authors use Surname1 **and** Surname2
  - More authors use Surname1 et al. (plural!)
  - Avoid configurations like:
    ... Dahl et al. [Dahl et al. 2007] ... (`\shortcite{}`)
  - By removing reference the sentence should still make sense

Illustrative Example

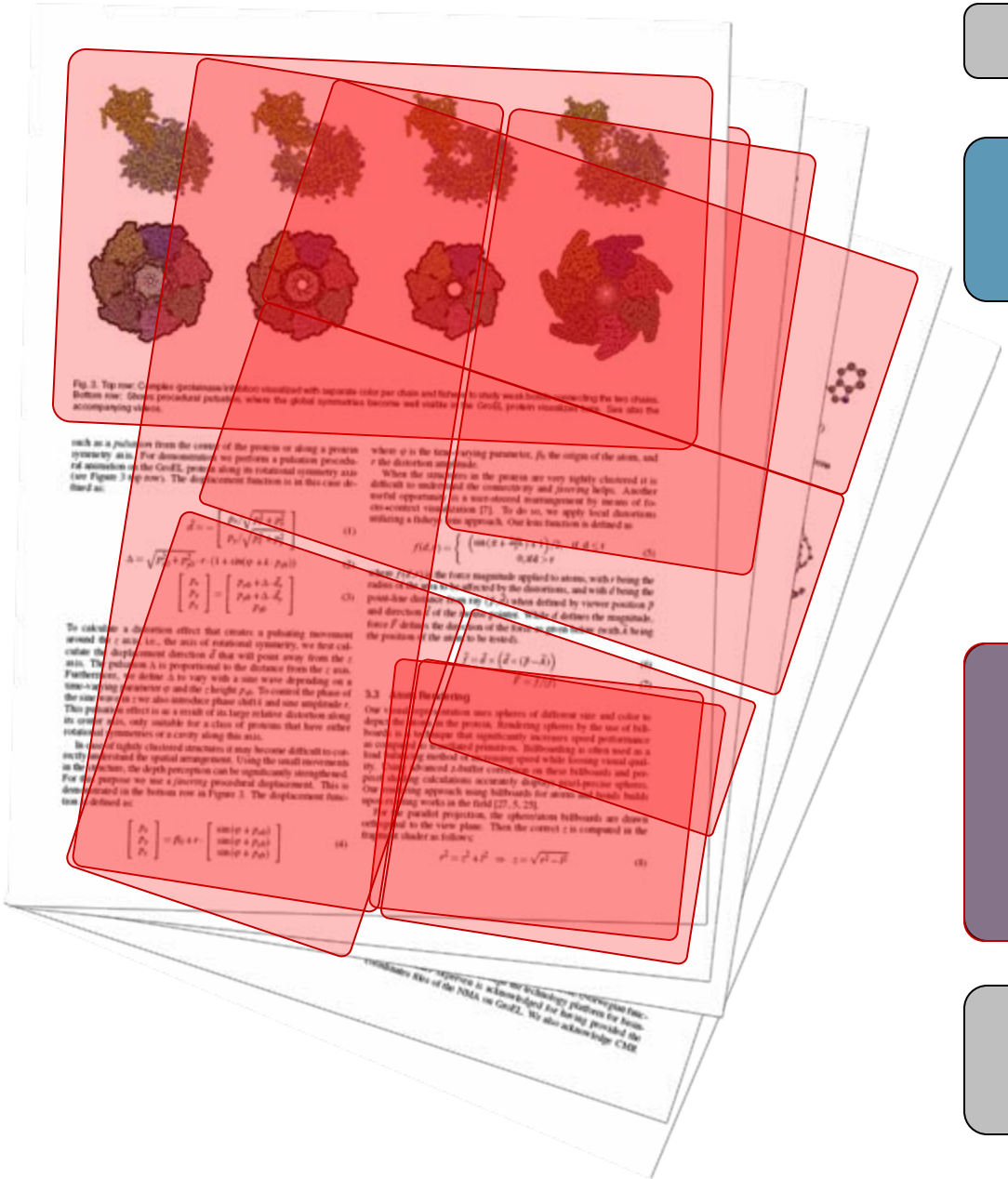


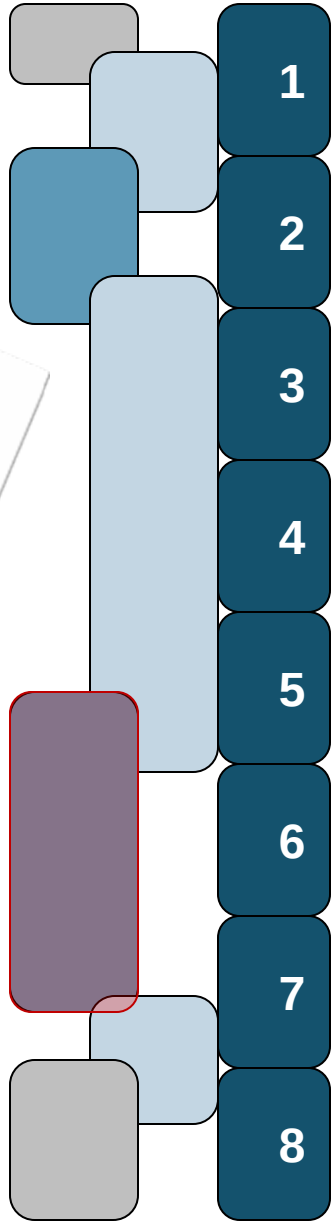
1
2
3
4
5
6
7
8

Main Sections

- The contribution of your work has to be presented as simple as possible
  - Start with the big picture
  - A workflow, pipeline, or framework graphics is very useful - visual table of contencts
  - Describe the algorithm in detail
  - Specific information on implementation
- Use illustrative graphics, tables, equations, pseudocode and discuss them in the main text
- Every statement should be either proven or referenced from past
- Reviewer has short time to evaluate, provide answers to her/his potential questions

Illustrative Example

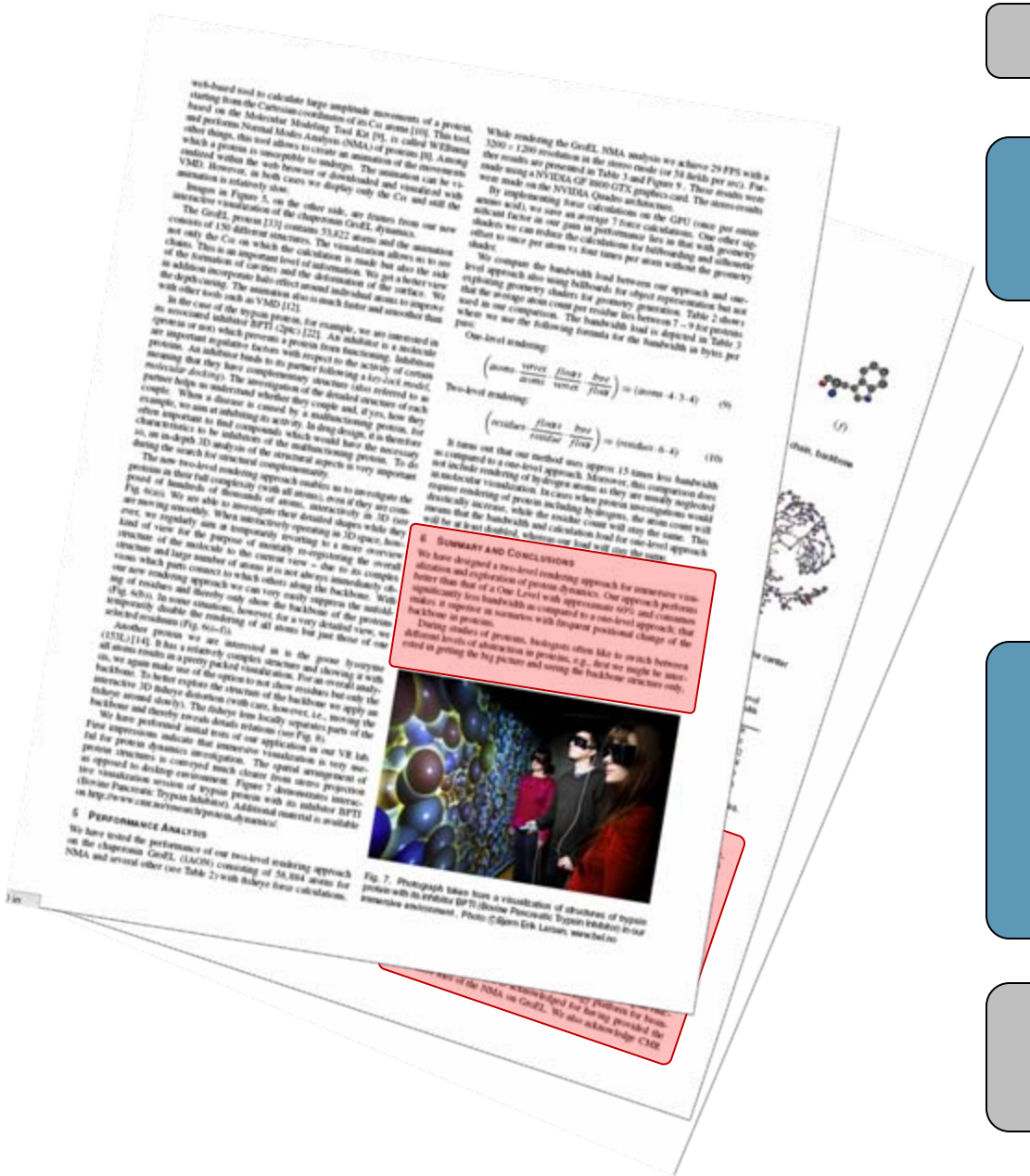




Validation and Results

- Demonstration of usefulness in the application domain
- Resulting rendered images especially important
  - Figure placement should be *close* to discussion addressing the figure
  - Optimally on the same page or double-page
- Performance analysis
- Convergence behaviour
- Data description
- Error analysis

Illustrative Example

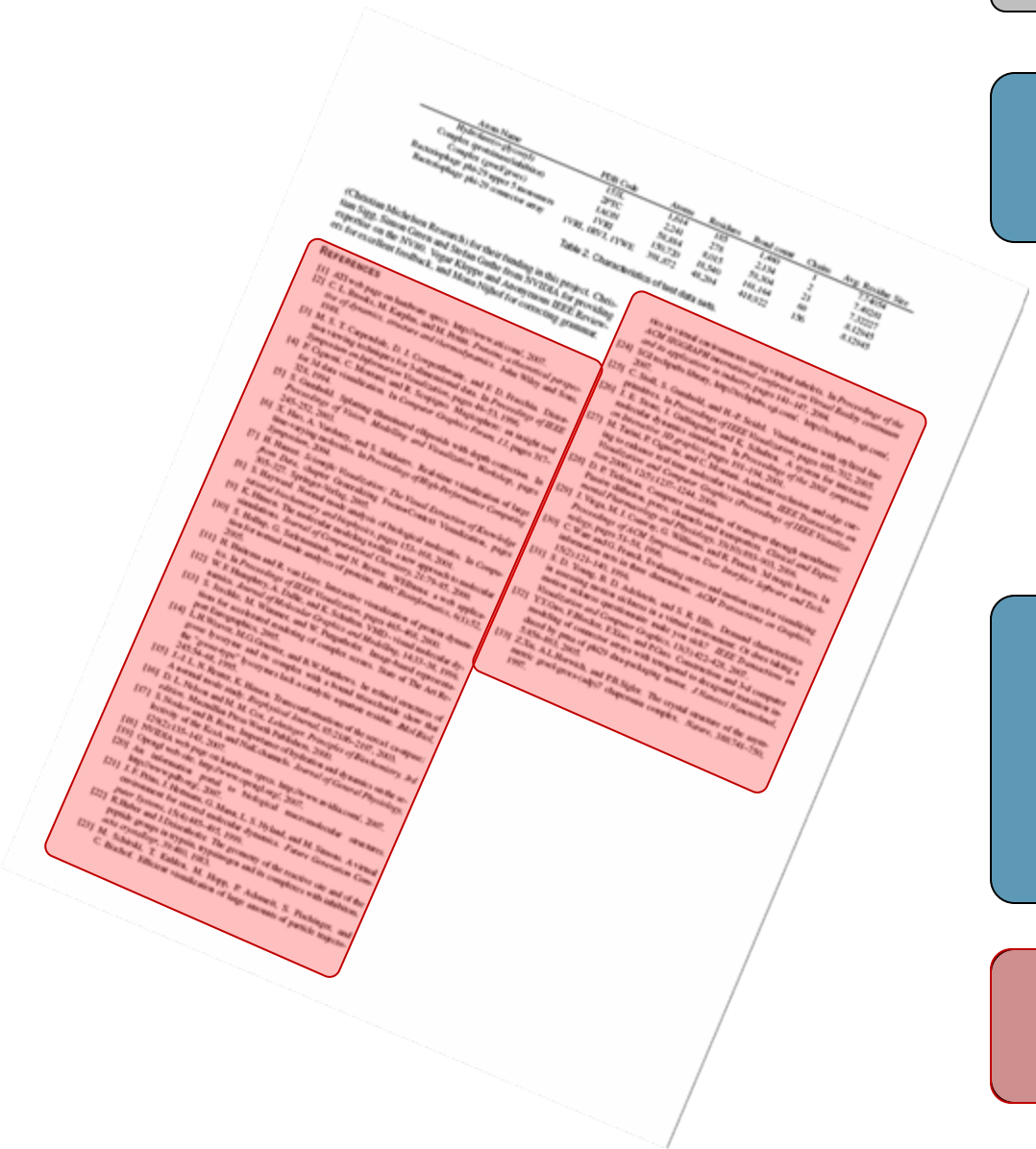


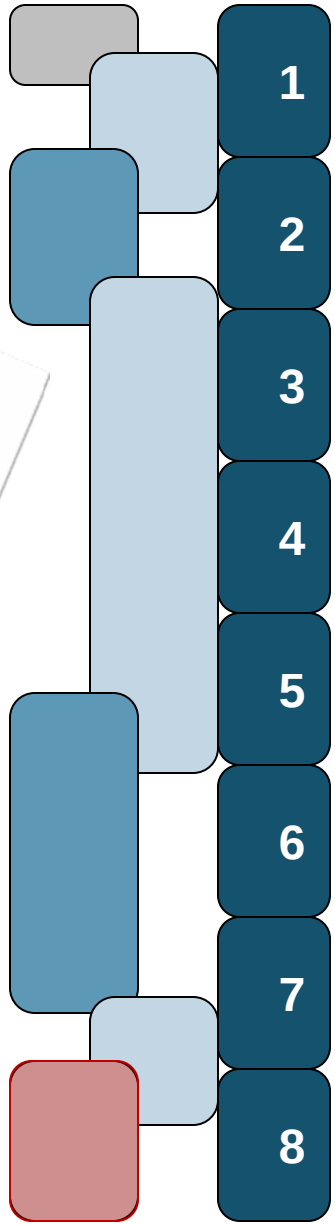
1
2
3
4
5
6
7
8

Summary, Conclusions, Outlook, Acks

- Few sentences on what has been achieved
- Conclusions is very important section!
  - What have you learned during the research
  - What should be people aware of
  - What turned-out to be most difficult
  - Under which assumptions the approach is valid
- Usually written in the past form
- Outlook
  - Further potential application areas
  - Improvements
  - Plans
- Acks: thx for data, illustrations, brainstorming

Illustrative Example





Thanks goes to

- Meister
- Helwig Hauser
- Thomas Theussl
- Werner Purgathofer
- Robert Tobler
- Dieter Schmalstieg
- Christian Breiteneder
- George Gopen and Judith Swan

Further Reading

- How to write a scientific paper
  http://www.cescg.org/guidelines/
- Scientific Writing (in German)
  http://www.ims.tuwien.ac.at/teaching/gma/nachlese/wiss_schreiben.php
- The Science of Scientific Writing
  http://www.amstat.org/publications/jcgs/sci.pdf
- Great page about giving a talk and writing a paper
  http://www.wit.at/events/peyton-jones/