# Writing in science:

how to strike a reader's interest
and convince them that you're right

# The role of writing in science

- Scientific results are not *real* results until published
- The peer review process is central to scientific integrity
- Scientific knowledge is largely established in the peer-reviewed literature
- Your success in science or industry is 1/3 writing ability (also: 1/3 oral communication and 1/3 technical ability)

# Venues for writing

- Abstracts
- Journal articles / papers
- Reports
- Proposals
- Books
- Notes

peer-reviewed,
latest results

What do you want to read?

# Think about who will read your work

- Majority will skim title and first sentences
- Some will read abstract and conclusions
- Fewer will skim through entire paper and figures
- Very small percentage will read thoroughly

- ***You want people to read, understand, acknowledge, and build upon your work. Design your writing to maximize these scenarios!***

# Four keys to writing: SASS principles

- **Structure** – provide an organized, clear picture of your work that involves a hierarchy of detail
- **Audience** – design your information and writing for the expected readers, not for yourself
- **Specific** – clearly delineate the problem being addressed and provide enough detail for someone to reproduce your results
- **Straightforward** – aim for maximum clarity and ease-of-understanding in your writing

# Structure

# The big pitfall

Serial, story-like writing =

- incomprehensible for quick readers
- poor advertising layout
- loss of main ideas

*"Building on the discussion in sections 1 and 2, we now find that equations 4 and 20 cannot be used in the general case."*

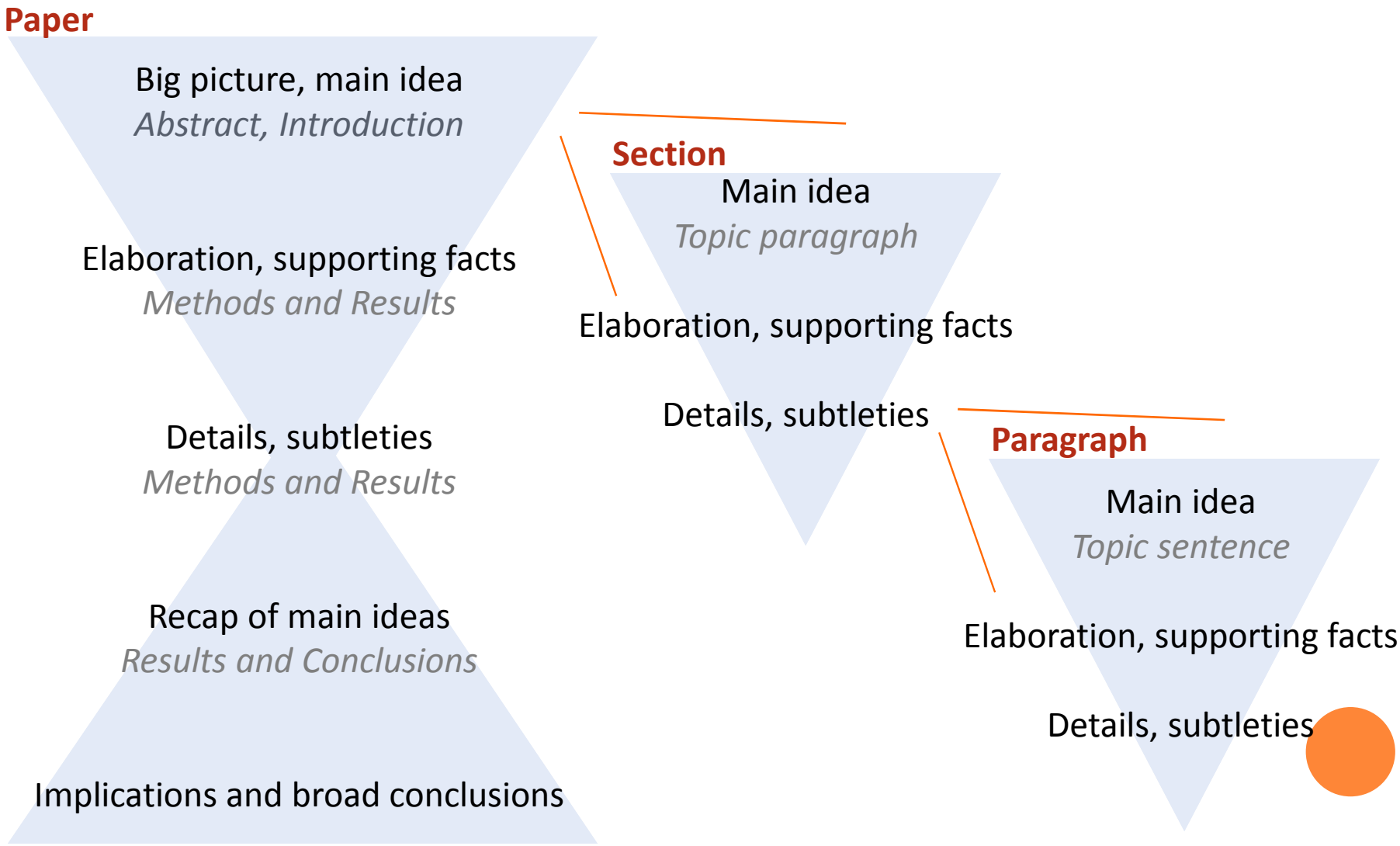
# A cardinal rule

Writing should be structured hierarchically at all levels, from main idea to details.

The details should be skippable while maintaining the big picture.

*"This procedure has limitations. The density- and temperature-dependence of the potential energy function preclude its use in certain kinds of simulations."*

# Hierarchy of ideas



# Design for reading in multiple ways

- **Big picture read**: quick skim of abstract and headings, subheadings
- **Main results read**: skim of results subheadings, figures, and conclusion
- **Methodology read:** detailed reading of methods
- **Quick read:** reading abstract and topic sentences
- **Full read:** sequential reading through the paper

# Basic structure of a scientific article

- **Title** – your tagline, advertising slogan, summary of everything
- **Abstract** – your chance to "hook" a reader
- **Introduction** – clearly define the problem and place your work in the context of the field; give the reader a big picture preview
- **Methods** – how you did it
- **Results** – what you did, what's surprising, how your work addresses the problem, how your results compare to others'
- **Conclusions** – summary of the problem and results; the broad implication of your results in the field; what's still unresolved

# Use descriptive subheadings

**Methods**

*A Database of Short Protein Fragment Simulations*

*Contact Metrics*

*Bayesian Classification Models*

*Training and Testing*

*Model Selection*

*Contact Prediction Success*

**Results**

*The Fragment Simulations Sample Around Native-Like Structures*

*Optimal Classification Models and Contact Metrics*

*Predicting Native Contacts and Conformations from Fragment Simulations*

*Extrapolating Inferences from Single Contacts to Larger Structures*

# The importance of topic sentences

Early efforts used Monte Carlo methods,[4-6] single-trajectory molecular dynamics,[7-12] and multi-trajectory molecular dynamics with massive distributed-computing,[13-15] to examine small peptide systems with well-defined secondary structures and proteinlike folding behavior.[16] …

*versus*

**Our work follows a long history of computational folding studies.** Early efforts used Monte Carlo methods,[4-6] single-trajectory molecular dynamics,[7-12] and multi-trajectory molecular dynamics with massive distributed-computing,[13-15] to examine small peptide systems with well-defined secondary structures and proteinlike folding behavior.[16] …

# How to draft a paper… fast

- Decide results and ideas you want to present. **Outline and arrange.**
- **Make figures and tables**.
- **Write the Methods and Results sections first**, then Conclusions, Introduction, and Abstract.
- Write text in an **iterative** fashion:
  - Start by writing a **topic sentence** for every paragraph .
  - Then, **bullet details and supporting ideas** underneath.
  - Convert to **rough body text** – don't perfect writing at this point.
  - Finally, edit the text to **streamline and focus** the writing.
- Add your **shorthand reference** notes as you go, then add actual references later in the very final stages.

# Abstract structure

- First sentence:
  **What problem are you talking about and why is it important?**

- 1-2 sentences:
  **What's been accomplished in the field?**
  **What remains to be solved?**

- 2-4 sentences:
  **What did you do?**
  **What are the main results that will be relevant to others' work?**

- Last sentence:
  **What's the big impact of your work?**

# Audience

# The big pitfall

Writing so as to impress yourself =

- too general, pedantic, sophomoric text
- unanticipated confusion
- readers miss what's important to them
- readers pissed off

*"Since the modern age, molecular simulations have been widely used in the simulation community."*

# A cardinal rule

Focus on what your audience will appreciate,
not on what you did or what basics you learned.

*"In the past two decades, simulations involving flat-histogram techniques have been widely used to study fluids, polymers, and biomolecules."*

# The hardest part: know your audience

- General scientific vs specialized audience → **journal-dependent**

- With regards to the audience,
  - **What's obvious?** *little discussion needed*
  - **What's a refresher of known facts?** *basic description & references*
  - **What details are relevant to your work?** *summary of prior studies*
  - **What's new?** *detailed but clear and organized elaboration*
  - **What's surprising?** *detailed arguments addressing potential questions and disbelief*

# Guidelines for references

- Clearly place your work in the **context** of what has been done
- **Cite everyone** – too many usually not a problem, but too few will irk reviewers and people in the field
- Never cite review articles or books when intending to prove a specific point – cite **original sources**
- **Compare** your results to others', but use good diplomacy:

*"Our method is superior to Shell's approach [4]."*

versus

*"Compared to the results in [4], our method has a rate of convergence that is 43% faster."*

# Specific

# The big pitfall

Poorly thought out phrasing and word choices =

- unfocused discussions
- unintended claims
- long, windy manuscripts
- opportunities for criticism

*"We show that our procedure produces models with more waterlike behavior."*

# A cardinal rule

Every word should have a purpose. If another word can be found that conveys more information about your work, use it instead.

*"We show that our iterative algorithm produces models that more closely replicate water's microscopic structure and bulk thermodynamic response functions."*

# Key points on which to be very specific

- What is **known** about the particular problem being studied?
- How does your work **differ** from what others have done in the field?
- What **methods** did you employ?
- What **problems** does your work solve?
- What are the **limitations** of your work?
- What **remains** to be determined?

# Do a specificity self-check on your text

- Could  it be confused to mean something else?
- Does it overstate the generality of your results?
- Does it state anything that you could not back up with a reference to a published work?
- Are references provided for general statements?

- ***Avoid long passages and discussions that carry little information.***
- ***Shorter manuscripts are harder to produce than longer ones because they require one to maximize information content.***

# Use power connecting words

- suggest, show, identify, signal, emerge, demonstrate (rather than “proves”)
- interpretation, development, paradigm, picture, argument
- couple, develop, integrate, construct, interface
- strategy, transferability, interplay, criteria, indicator
- quantify, assess, detect, address, identify, distinguish
- systematic, preliminary, underlying, conspicuous, marginal, consistent
- unify, broaden, enhance, distort, exceed, contribute
- amenable, initiated, critical

# Straightforward

# The big pitfall

Pontification and flowery language =

- confused readers
- bored readers
- no readers

*"Thinking about the trend in Figure 5, but keeping in mind the special cases mentioned earlier, it is partially apparent that the system has a first order transition, although we do not perform a detailed analysis of fluctuations."*

# A cardinal rule

What would you want to read?

- Long passages with creative text?
- Short, to-the-point discussions?

*"Figure 5 suggests that the system has a first order transition. Still, we do not perform a fluctuation analysis and one must keep in mind the special cases mentioned earlier."*

# Simplify grammar

- **Watch out for too many commas and parenthetical phrases**

*The density of states, that is to say the degeneracy of energy levels in a system, is related to the entropy, hence can be used to calculate all thermodynamic properties, through special simulation techniques."*

versus

*"The computed density of states provides the entropy and can be used to calculate all thermodynamic properties of a system."*

- **Rearrange awkward phrases to be more direct**

*"That liquid water is vitally important in virtually all aspects of our lives is uncontested."*

versus

*"Liquid water is vitally important in virtually all aspects of our lives."*

# Simplify grammar

- **Generally OK to use first person**

*The sample was incubated for 30 minutes and subsequently left until it reached a temperature of 400 K."*

versus

*"We incubated the sample for 30 minutes and waited until it reached 400 K."*

- **Split long, convoluted sentences into multiple ones**

*"Generally, although there has been little direct evidence except in a few cases, it is recognized that folding proceeds according to a collapse process, starting first in an unfolded ensemble and proceeding through a single intermediate."*

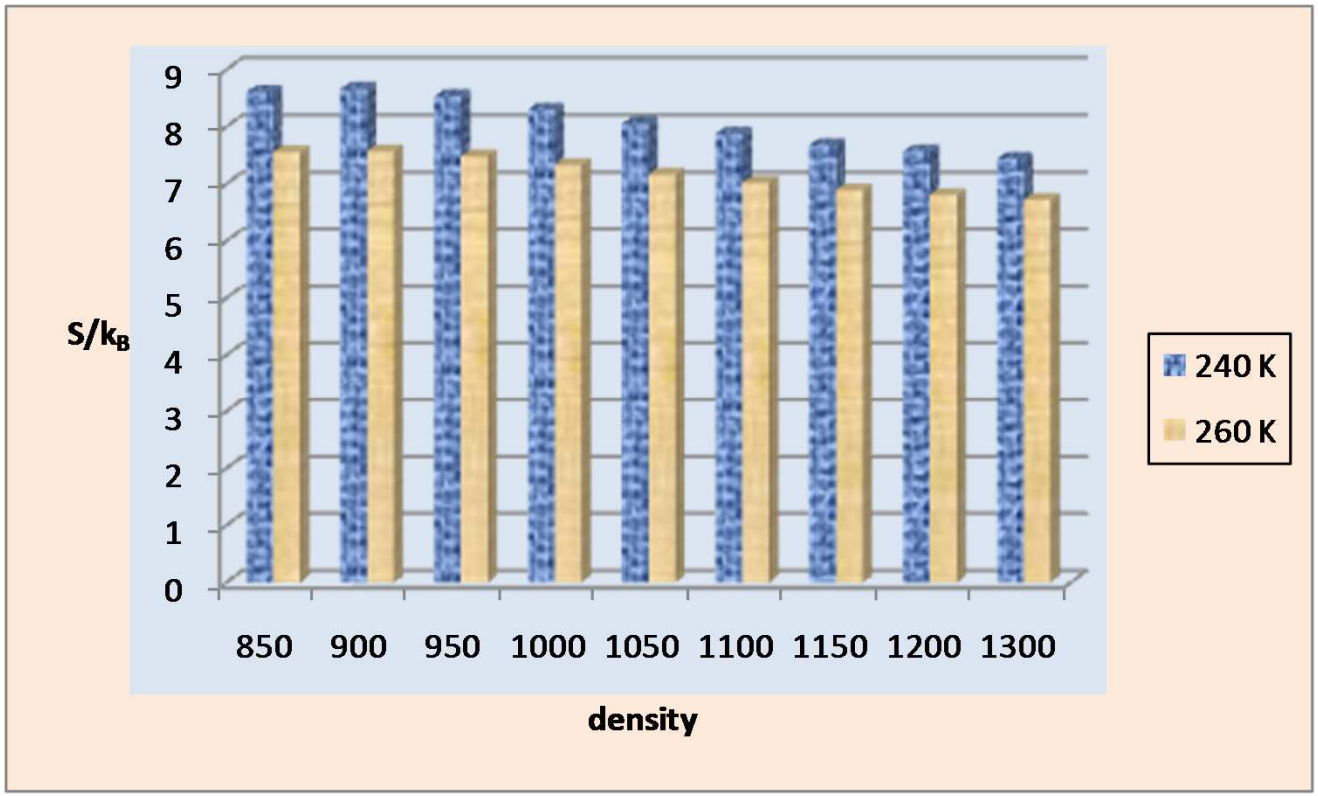
versus

*"Folding is generally thought to proceed according to a collapse process, even though there has been little direct evidence. This process starts in an unfolded ensemble and proceeds through a single intermediate."*

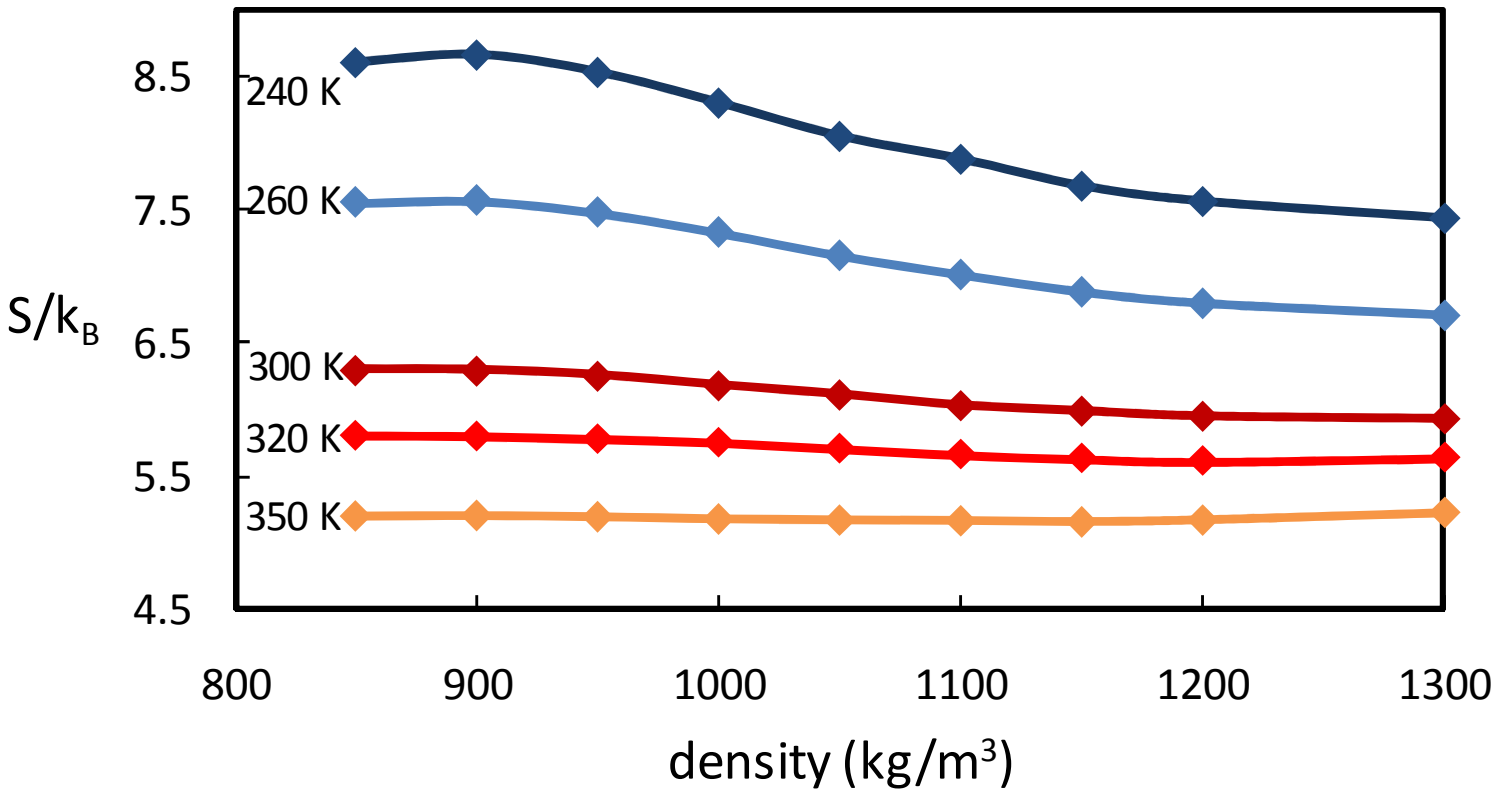
# Presenting data in figures

- Remove extraneous marks!
- Make sure figures are legible at printed resolution (3.25 inch width journal standard)
- Avoid fancy coloring and design – all colors should facilitate interpretation of the data
- Use fit lines or series lines to guide the eye
- Avoid putting too much on a single graph or figure
- Use informative captions!
- ***Strive to make the figure as easily digestible as possible for the reader***

# Horrible


9
8
7
6
5
4
3
2
1
0
S/k_B
850
900
950
1000
1050
1100
1150
1200
1300
density
240 K
260 K

# Good


8.5
7.5
6.5
5.5
4.5
S/k_B
240 K
260 K
300 K
320 K
350 K
800
900
1000
1100
1200
1300
density (kg/m³)

# The Triple Challenge

**Write a three-sentence summary of your most recent project.**

You should convey the import, basic idea, and main results of your work in a way that will stimulate further interest from the audience.