Bioinformatics II
Winter Term 2016/17

universitätbonn
Rheinische
Friedrich-Wilhelms-
Universität Bonn

Chapter 10:
Diffusion MRI

Jun.-Prof. Dr.-Ing. Thomas Schultz

URL: http://cg.cs.uni-bonn.de/iaan/
E-Mail: schultz@cs.uni-bonn.de
Office: Friedrich-Ebert-Allee 144, 53113 Bonn

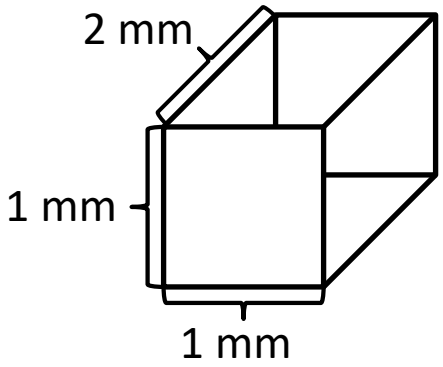
January 31, 2017

10.1 Diffusion Tensor MRI

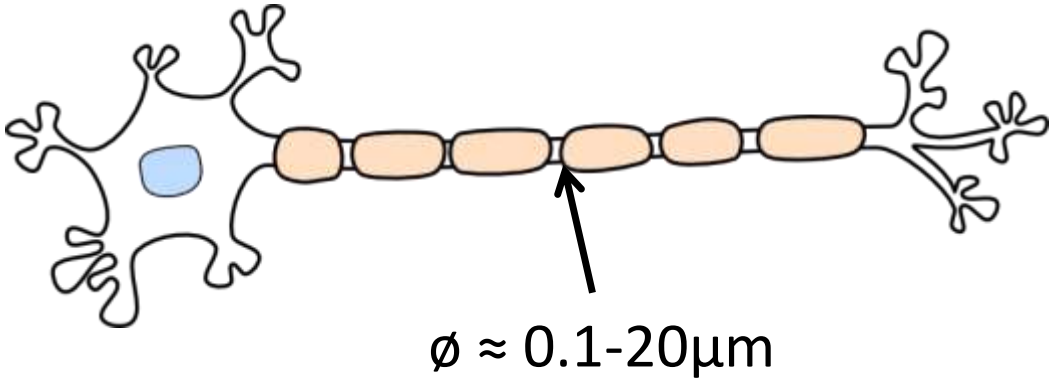
Introduction to Diffusion MRI

Goal: Investigate the microstructure of biological tissue using Magnetic Resonance Imaging (MRI)

Challenge: Voxel size is far too large to resolve the structures of interest



Voxel size



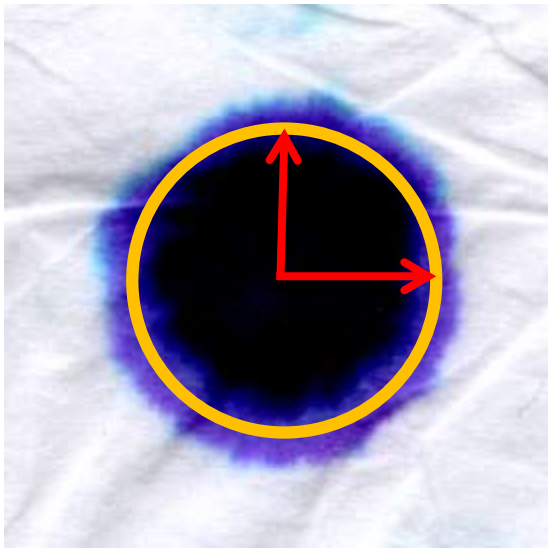
Axon (nerve fiber) size

Introduction to Diffusion MRI

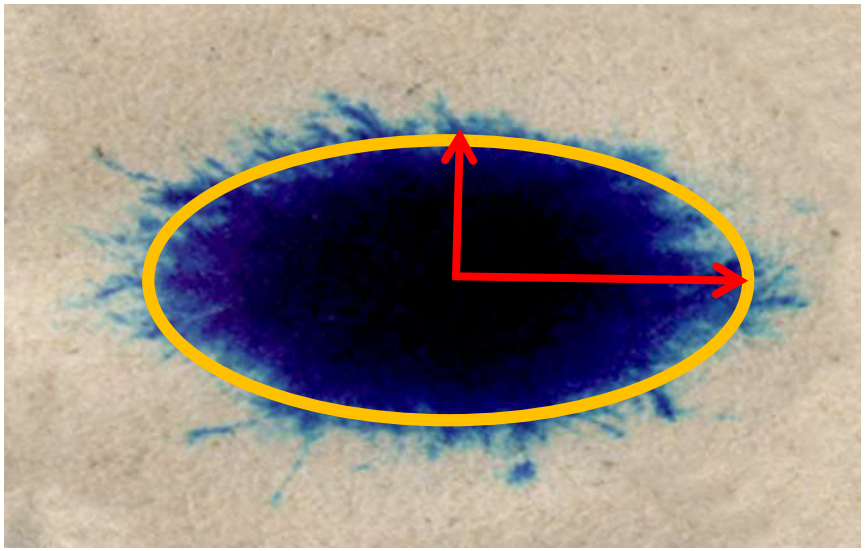
Approach: Use water molecules as a contrast agent

- Exploits their spontaneous heat motion at the desired spatial scale

Analogy: Observe diffusion of ink on paper



Kleenex



Newspaper

Images from Gordon Kindlmann

Brownian Motion and Diffusivities

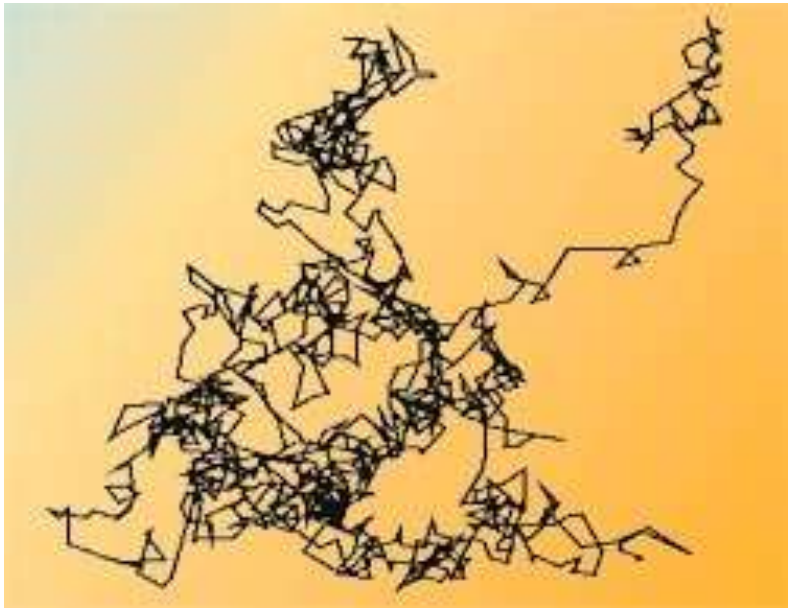
Molecular diffusion = Brownian motion of water molecules

- Random walk due to thermal energy

- Measured with the Diffusion Coefficient *D*
  - Proportionality constant in Fick's law:

$$\boldsymbol{J} = -D\nabla \mathrm{c}$$

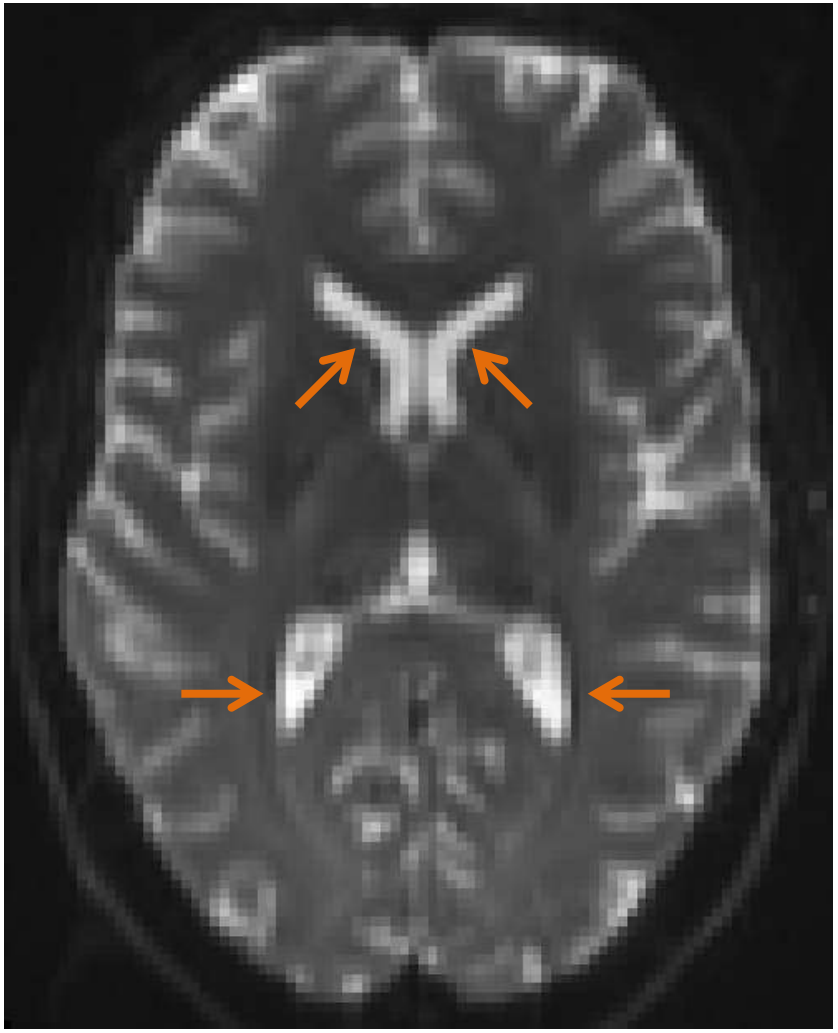
  - Brownian motion happens even without a concentration gradient („self-diffusion“)



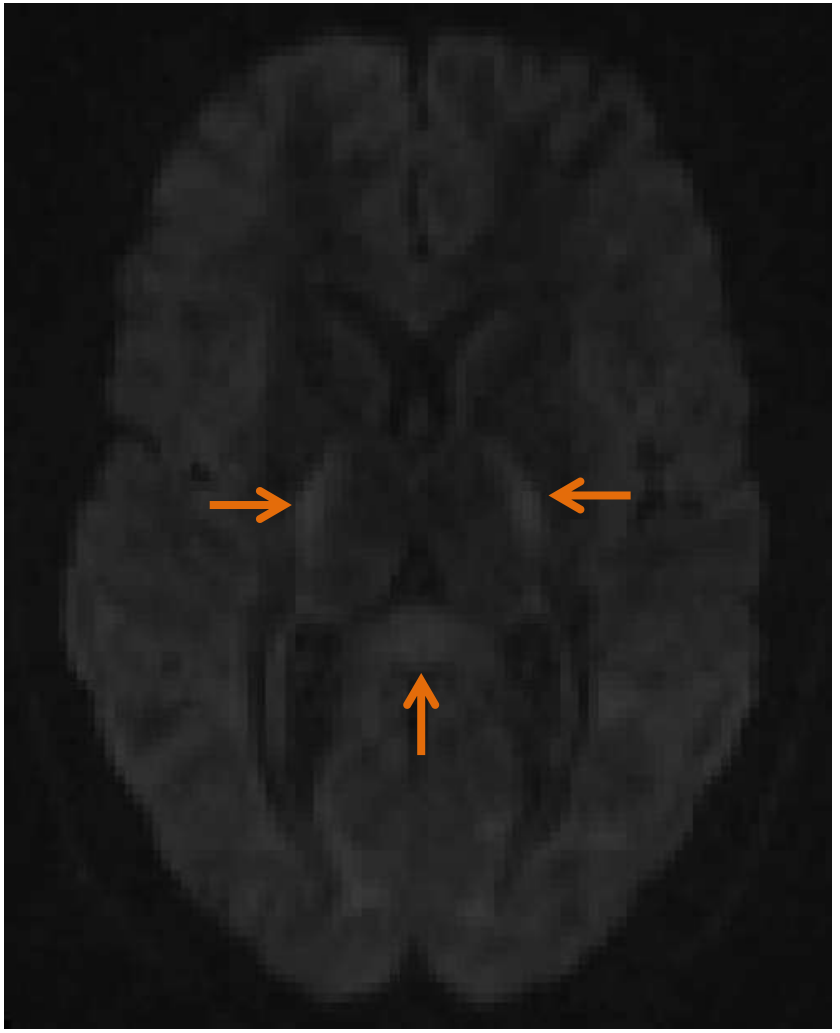
- In white matter: $D \approx 1\text{-}5 \cdot 10^{-4}$ mm^2/s

- Due to interactions with cellular structures, *D* depends on spatial direction (anisotropic diffusion)

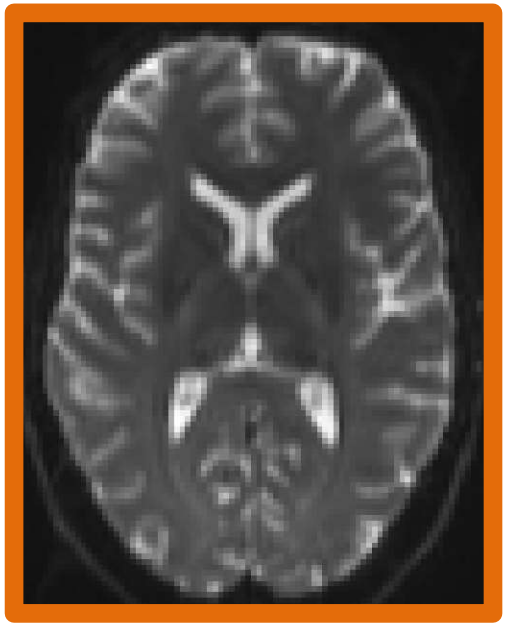
Introduction to Diffusion Weighting

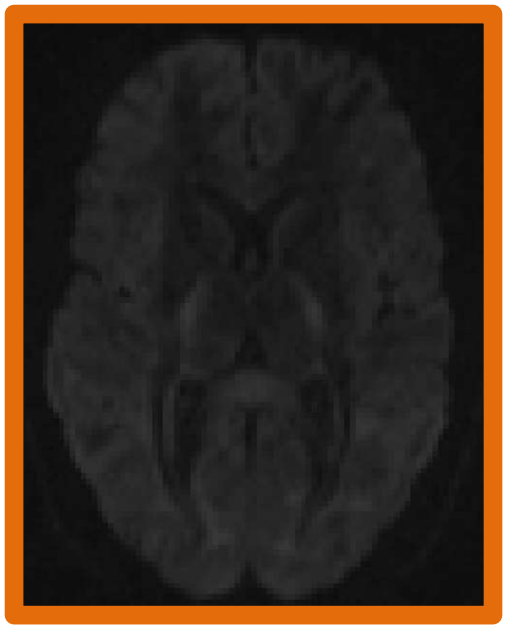


Standard T_2 MRI



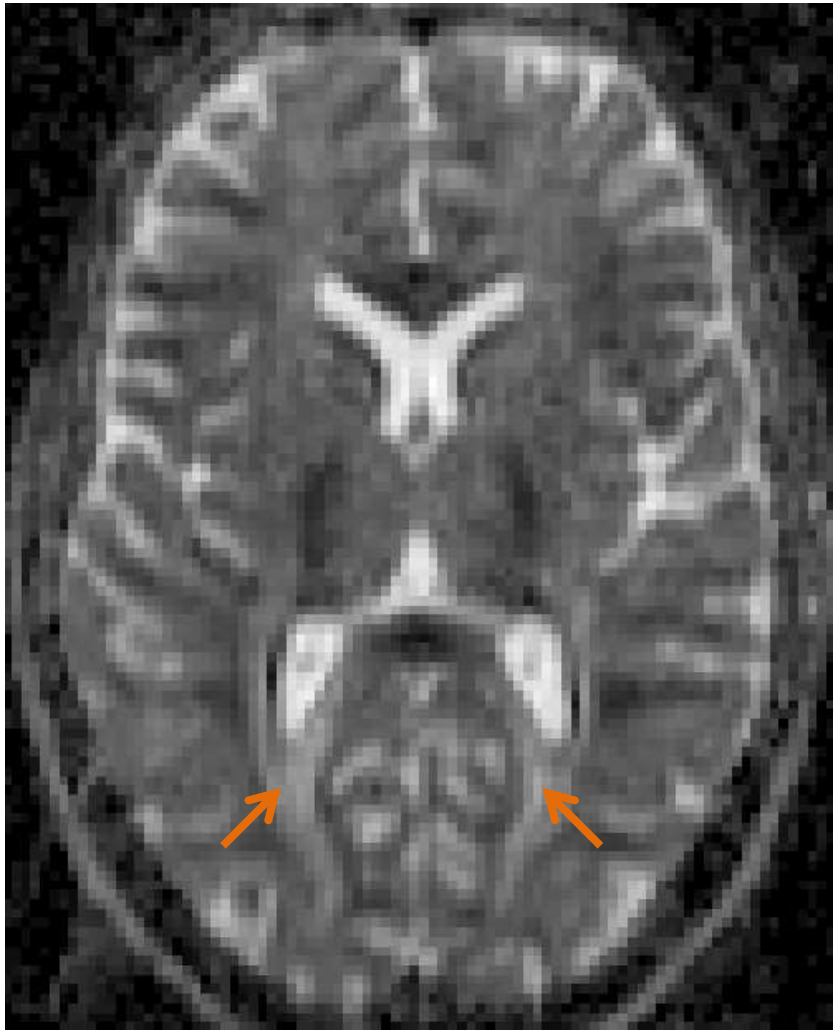
Diffusion Weighted MRI



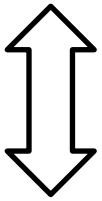


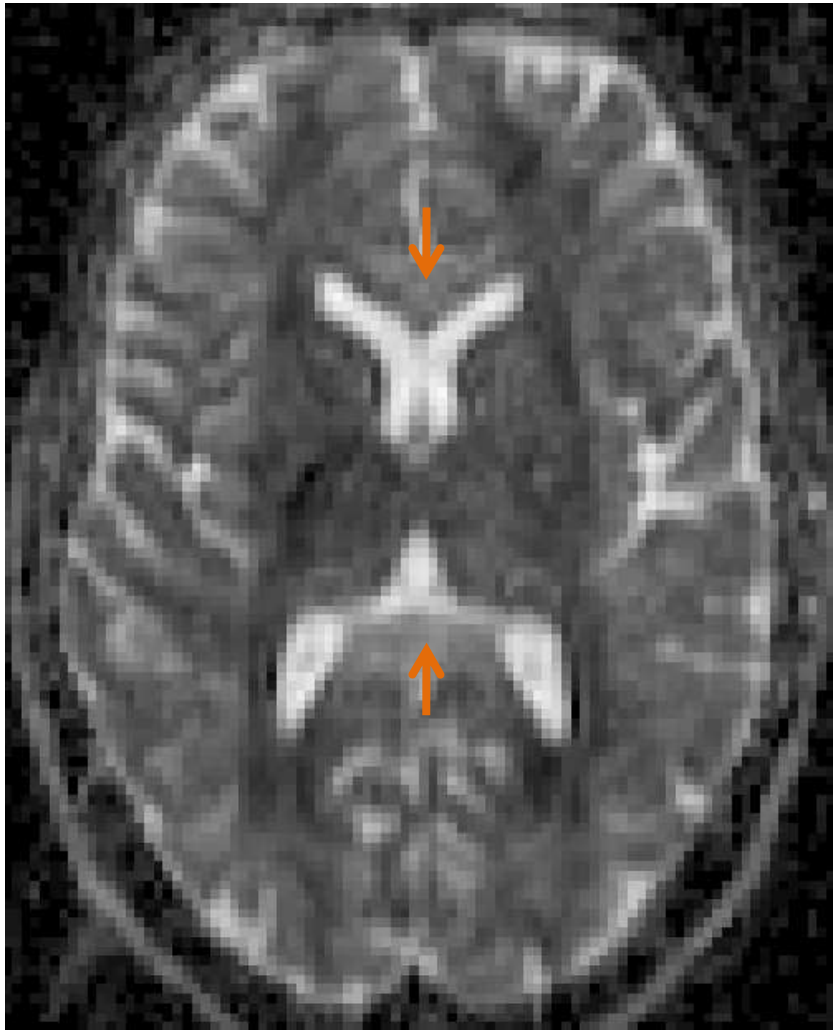
Stejskal-Tanner Equation:

$$S(d) = S_0 e^{-bd}$$

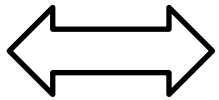


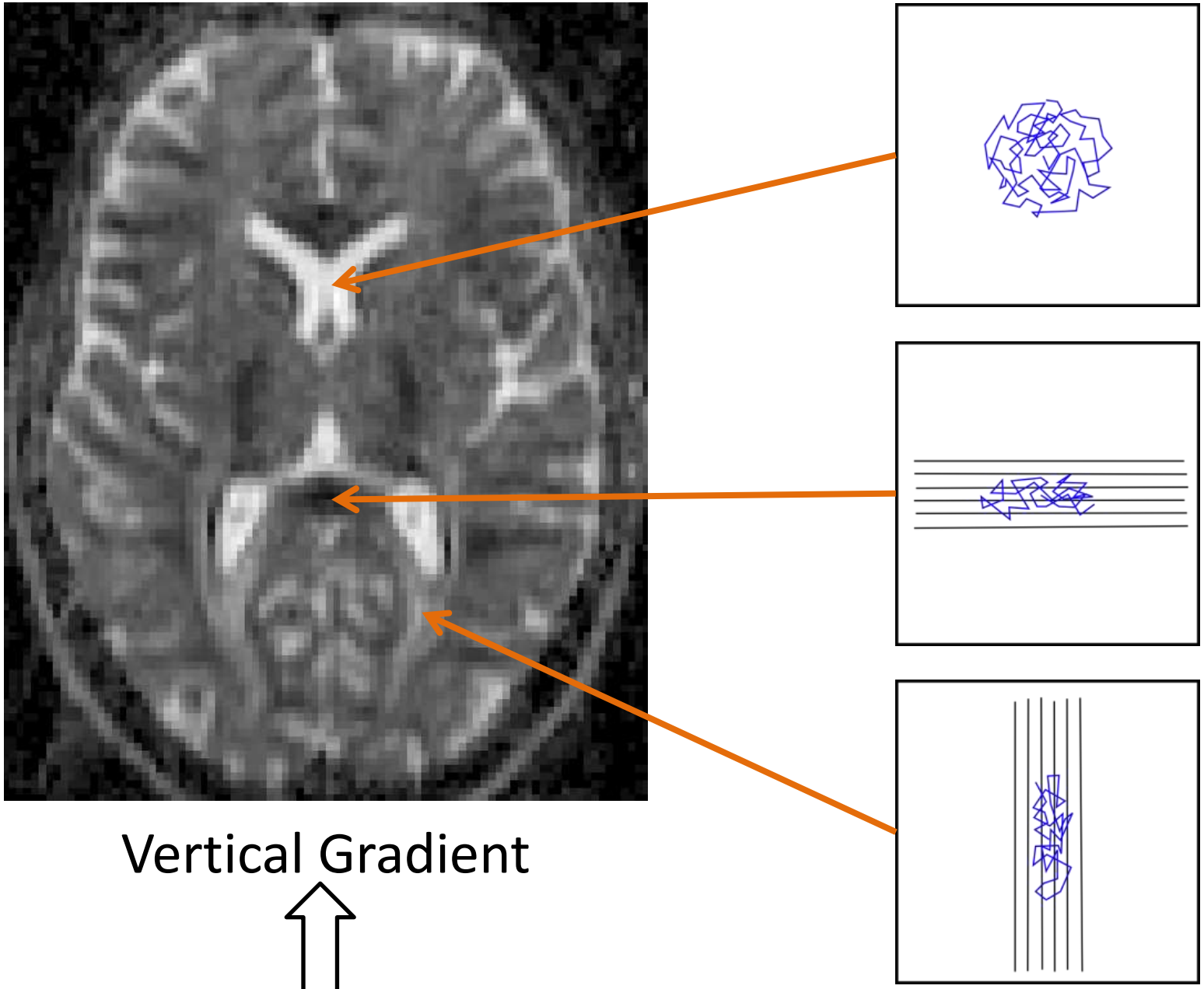
Vertical Gradient



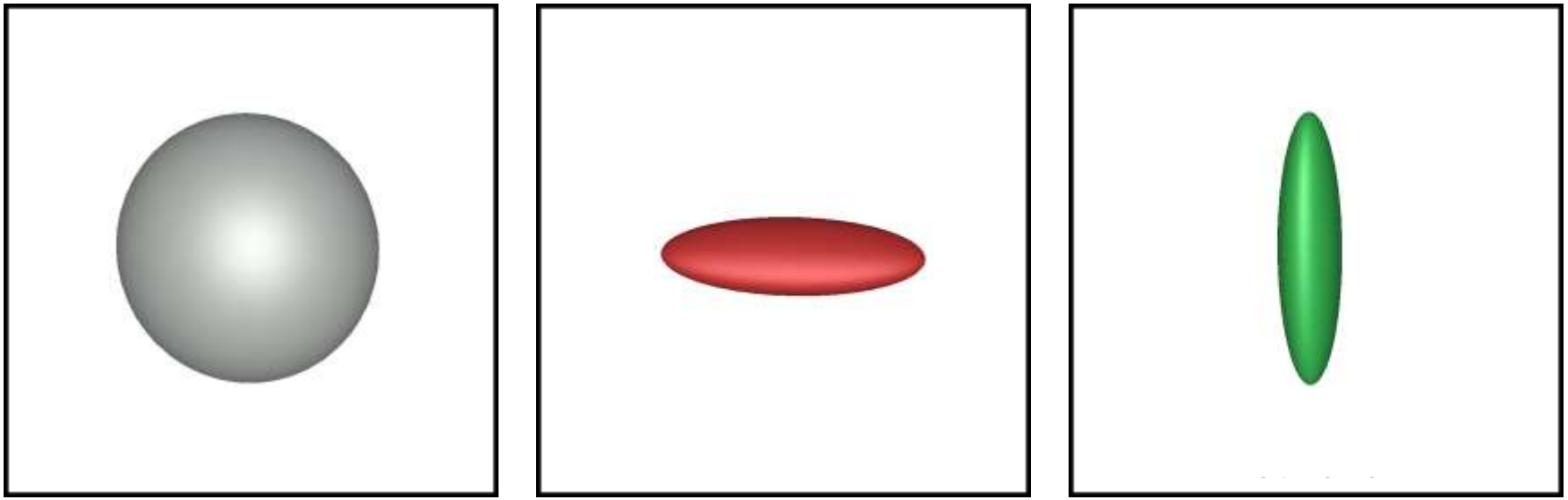


Horizontal Gradient




Vertical Gradient

The Diffusion Tensor Model



$$S(D(\mathbf{x})) = S_0 \mathrm{e}^{-bD(\mathbf{x})}$$

$$D(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{D}\mathbf{x}$$

D = 3x3 symmetric matrix

"Diffusion Tensor"

Eigenvector Decomposition

Diffusion tensor **D** decomposes into:

- 3 eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$
  - If **D** is positive definite: All $\lambda > 0$
  - *Note:* Diffusivity is a non-negative physical quantity. Due to measurement noise, we might obtain $S_0 < S(d)$, which can lead to **D** with negative λ's

- 3 orthogonal eigenvectors $\mathbf{e}_1/\mathbf{e}_2/\mathbf{e}_3$
  - $\mathbf{e}_1$ indicates single main fiber orientation

Axes of **tensor ellipsoid** are

- aligned with eigenvectors and
- scaled by eigenvalues

Estimating Diffusion Tensors

- **Stejskal-Tanner Equation:**

$$S(\mathbf{x}) = S_0 \mathrm{e}^{-b\mathbf{x}^T\mathbf{D}\mathbf{x}}$$

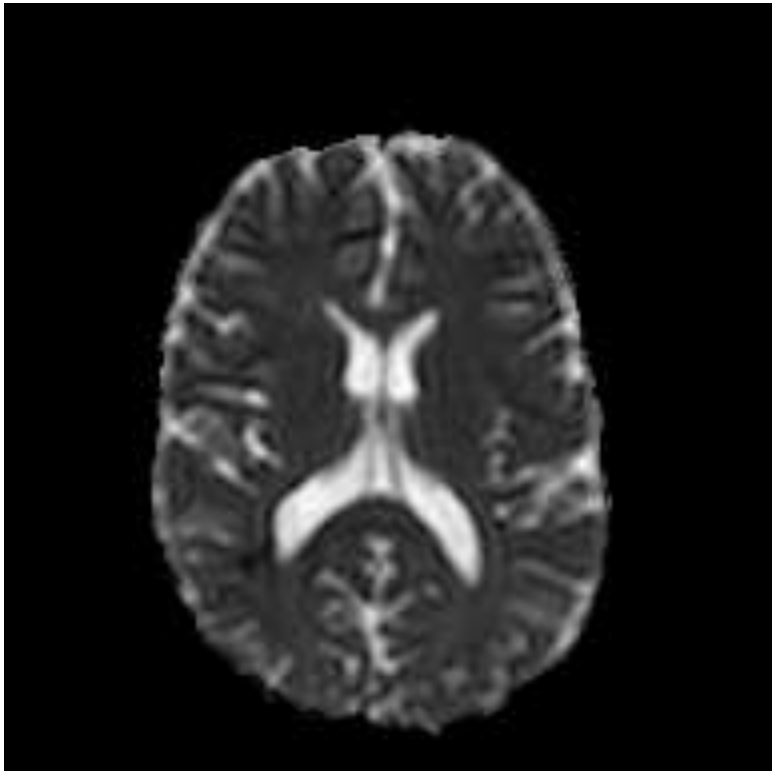
- Taking the logarithm and solving for **D** produces a system of linear equations (one per measurement *i*):

$$\sum_{k,l} \underbrace{\mathbf{x}_i^T\mathbf{D}\mathbf{x}_i}_{[\mathbf{D}]_{kl}[\mathbf{x}_i]_k[\mathbf{x}_i]_l} = -\frac{1}{b}\ln\frac{S(\mathbf{x}_i)}{S_0}$$

  - **D** = symmetric 3x3 matrix → six free variables
    - Due to noise, we usually take more than six measurements and find least squares solution

Matrix Trace / Mean Diffusivity

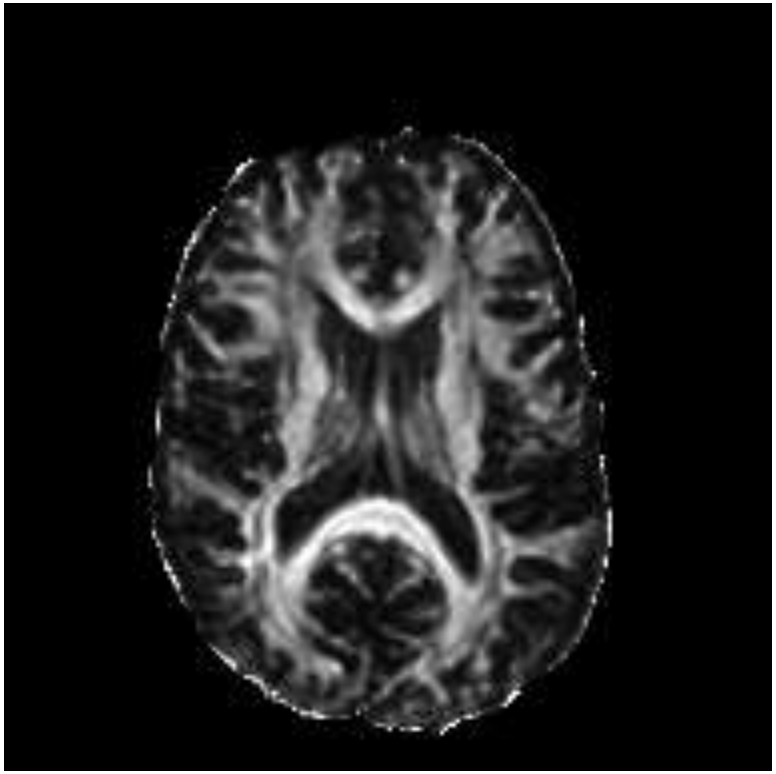
- **Matrix Trace**
  - Sum of diagonal elements: $\mathrm{tr}(\mathbf{D}) = D_{xx}+D_{yy}+D_{zz}$
  - Same as sum of eigenvalues: $\lambda_1+\lambda_2+\lambda_3$



- **Mean Diffusivity (MD)**
  - Average: $\mathrm{tr}(\mathbf{D})/3$
  - Interpretation in DT-MRI: Average diffusivity (over all directions)
    - Approximately constant in healthy tissue
    - Sensitive to edema, necrosis

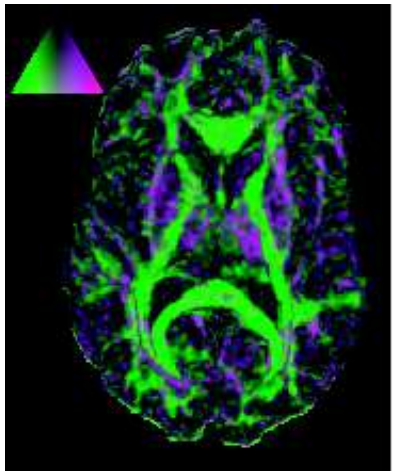
Fractional Anisotropy (FA)

- **Fractional Anisotropy**
  - Quantifies the degree of anisotropy
    - Correlates with fiber density / integrity
    - Also correlates with orientation dispersion
  - Based on variance of eigenvalues
  - Normalized to [0,1] (if positive definite)



$$\mathrm{FA} = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - \mathrm{MD})^2 + (\lambda_2 - \mathrm{MD})^2 + (\lambda_3 - \mathrm{MD})^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}$$

Westin Measures

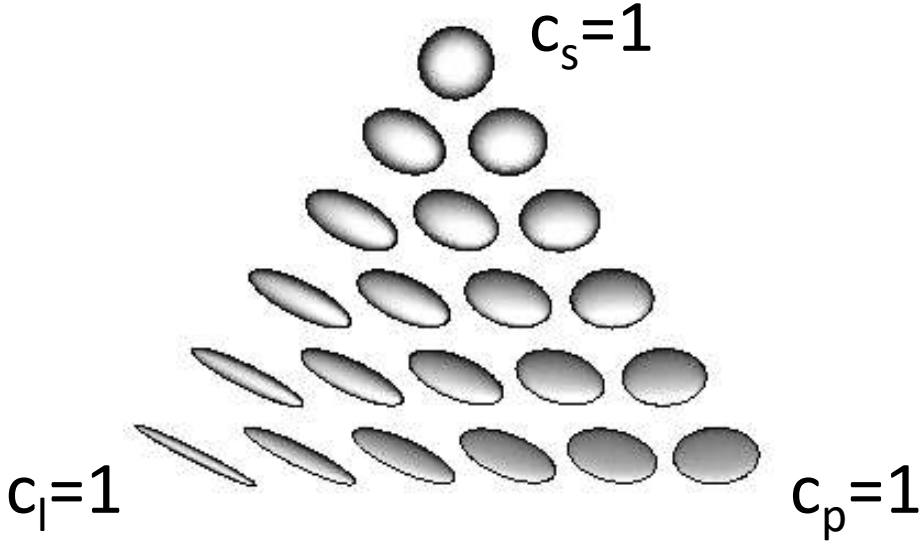


- Westin's c_l, c_p, c_s quantify the extent to which the tensor ellipsoid is linear / planar / spherical

$$c_l = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3}$$

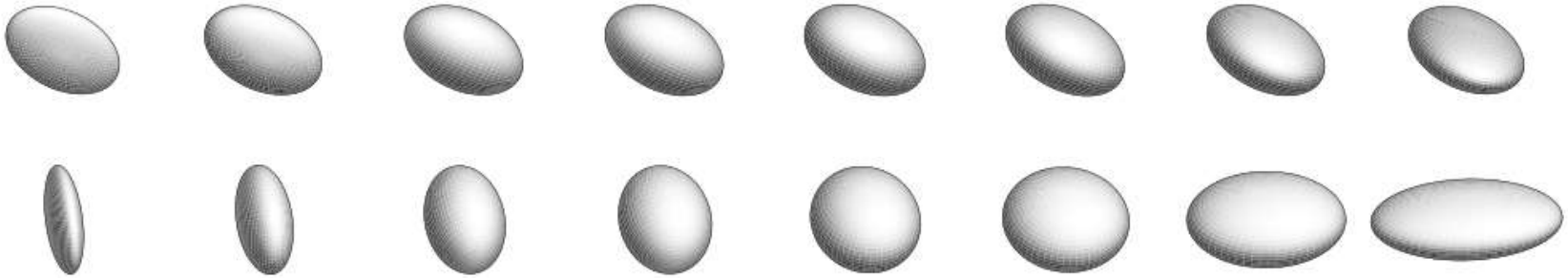
$$c_p = \frac{2(\lambda_2 - \lambda_3)}{\lambda_1 + \lambda_2 + \lambda_3}$$

$$c_s = \frac{3\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}$$

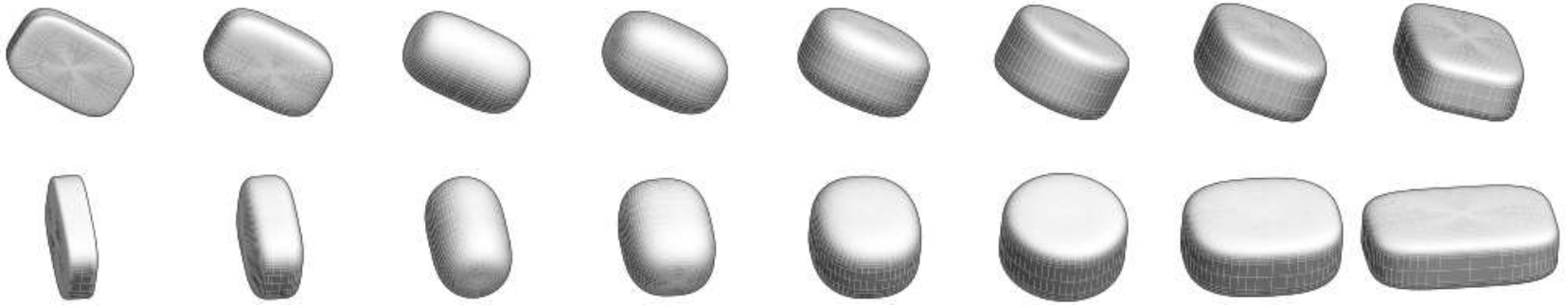


Superquadric Tensor Glyphs

- Ellipsoids suffer from **visual ambiguities:**



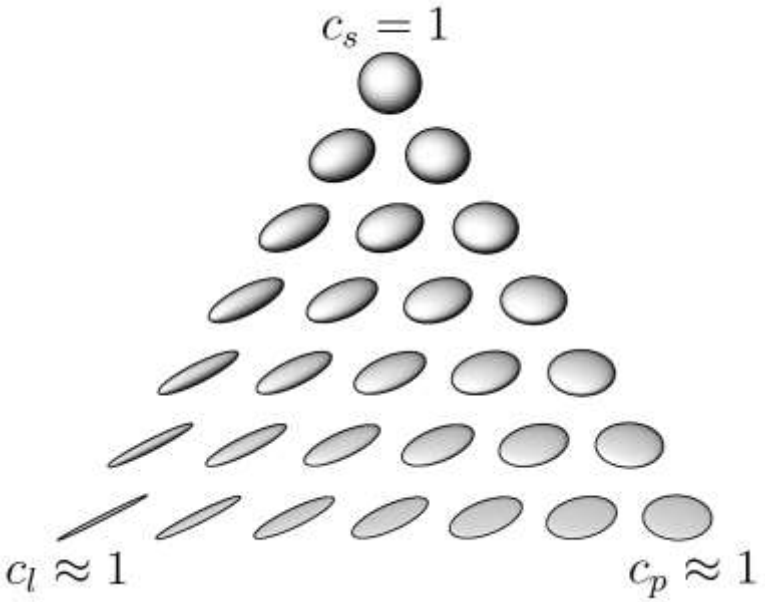
- **Superquadric Glyphs** greatly reduce them:



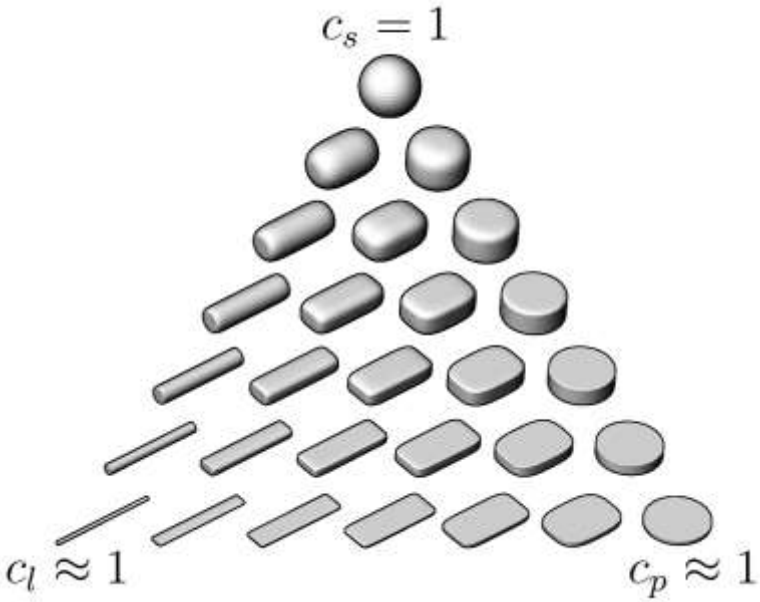
Images taken from Kindlmann [2004]

The Idea Behind Superquadric Glyphs

- **Ellipsoids** are transformations of the sphere
- **Superquadrics** smoothly interpolate between sphere, cylinder, and box



Ellipsoids



Superquadrics

Summary: Diffusion Tensor MRI

- DT-MRI is based on measuring magnitude of **Brownian heat motion**
  - Unlike standard MRI, DT-MRI is **quantitative** (calibrated), at least in theory
- **Probe for tissue microstructure:** Interactions between water and tissue leave a "footprint" on diffusion behavior
  - Exploring it requires many measurements, 5 minutes or more scan time for full-brain analysis
- **Model:** Exponential decay with quadratic diffusivity
  - Diffusion Tensor = Coefficients of the quadratic form
  - Can be thought of in terms of diffusion ellipsoid
  - Main parameters for statistical analysis: Fractional Anisotropy (FA), Mean Diffusivity (MD)

10.2 Tract-Based Spatial Statistics

Skeletonization: Why?

- **Obvious approach** to full-brain analysis of Fractional Anisotropy maps is "VBM-Style": Normalization + Smoothing + Statistical Mapping
  (no segmentation or modulation needed)
- **Problems:**
  - Would like to avoid smoothing, since it works against our goal of breaking the resolution limit of MRI
  - If alignment is imperfect, results are difficult to interpret

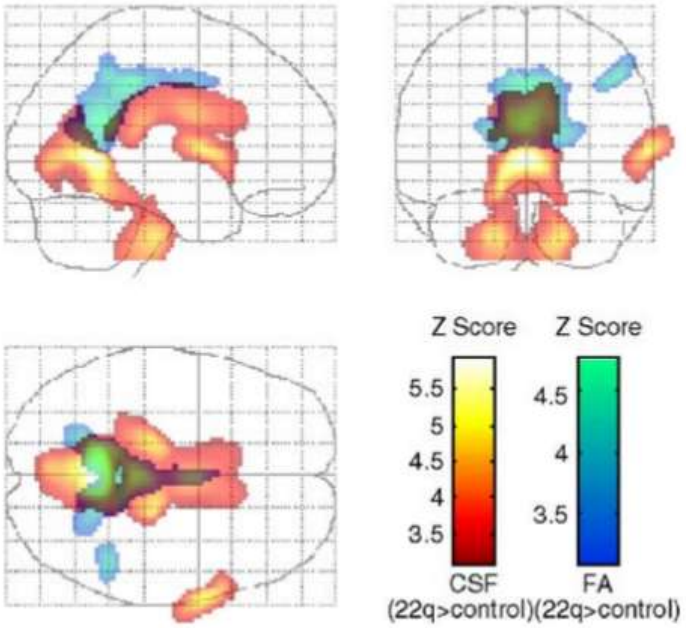


Image from [Simon et al. 2005]

Tract-Based Spatial Statistics (TBSS)

- **Idea:** Limit statistical analysis to the center of white matter structures
  - Fractional Anisotropy highest near the center
  - Given the “mean white matter skeleton” as a surface, we can project largest FA values from all subjects to it
  - Reduces number of tests and compensates inaccuracies in the initial registration without any smoothing

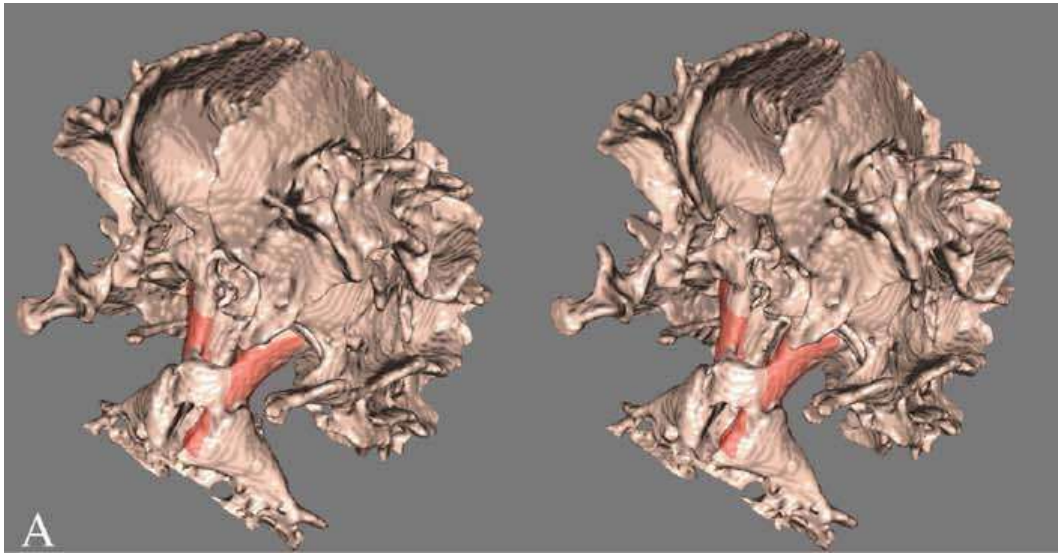


Image from [Smith et al. 2006]

The TBSS Pipeline

Fit Diffusion Tensor Model → Compute FA Maps

↓

Nonlinear Normalization → Study-Specific Template

↓

Skeletonize Mean FA Map → Project Individual FA Maps

↓

Statistical Analysis, Restricted to the Skeleton

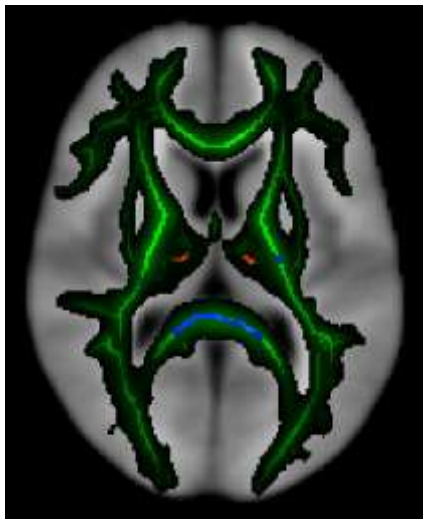


Image from Oxford's fMRIB lab

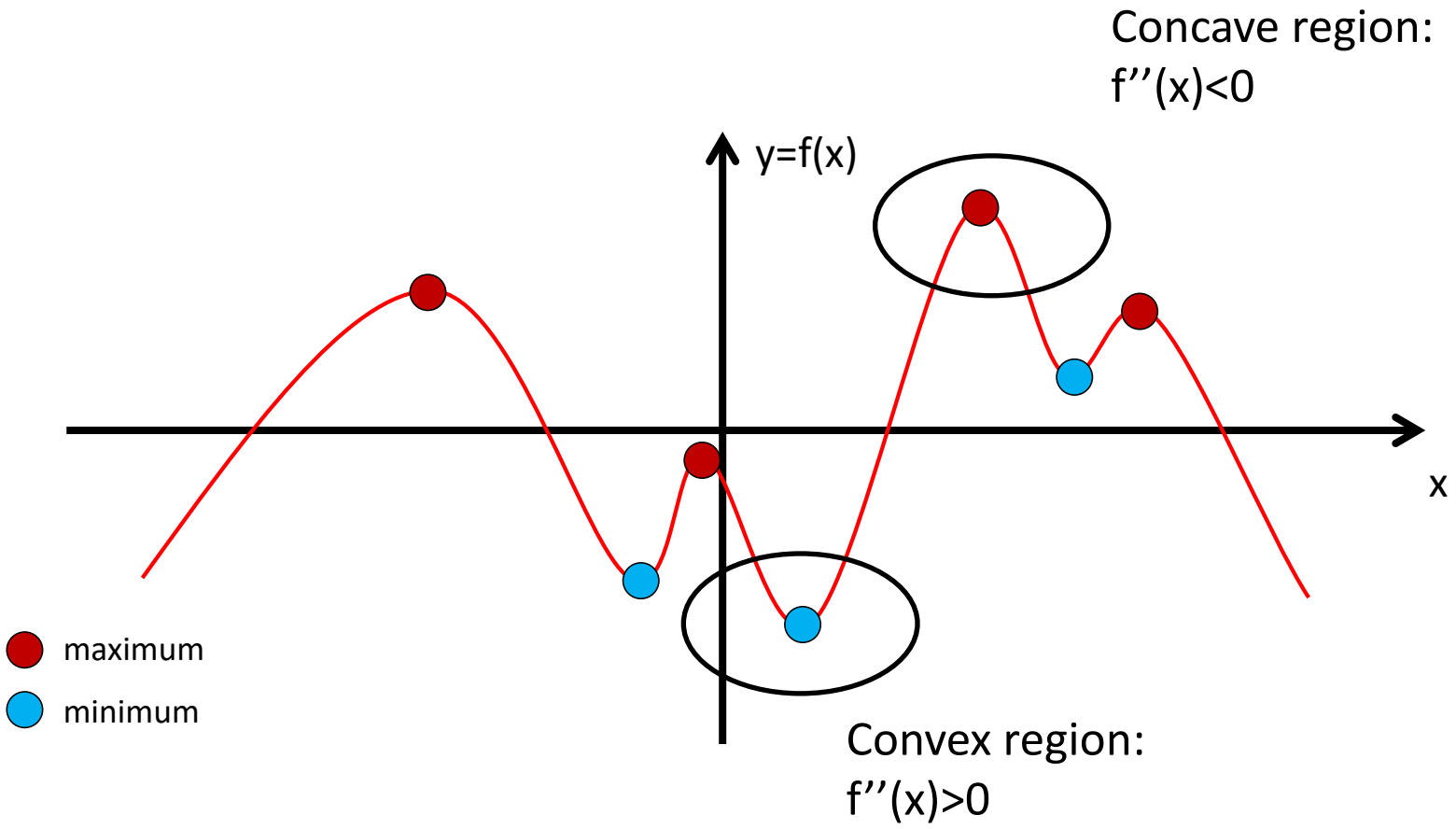
Intuition Behind Ridge Lines

Definitions by Merriam-Webster:

- ridge = “an elongate crest”
- crest = “top line of mountain or hill”

Image: Wikipedia

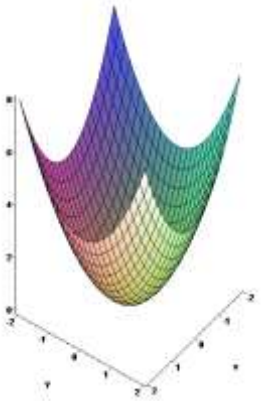
Reminder: Curvature in 1D



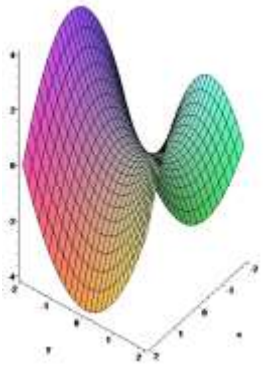
Reminder: Curvature in 2D

Surfaces have two *principal curvatures* whose signs may or may not agree.

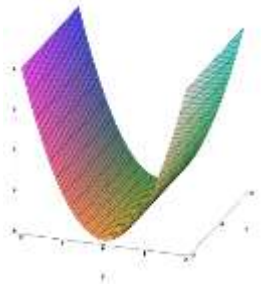
Examples:



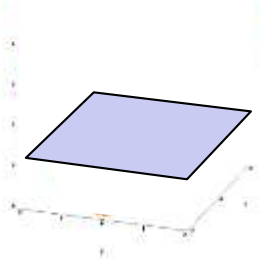
$\kappa_i > 0$



$\kappa_0 > 0, \kappa_1 < 0$



$\kappa_0 = 0, \kappa_1 > 0$



$\kappa_0 = 0, \kappa_1 = 0$

...

Eigenvectors and -values of Hessian Matrix

- *Reminder:* The Hessian matrix is the matrix of partial second derivatives:

$$H = \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{bmatrix}$$

- Since Hessian matrix is real and symmetric (for sufficiently smooth *f*),
  - its eigenvectors $\mathbf{e}_i$ are orthogonal
    - can serve as a basis of the domain
  - its eigenvalues λ_i are real
    - $\lambda_i < 0$: *f* is concave in direction $\mathbf{e}_i$
    - $\lambda_i > 0$: *f* is convex in direction $\mathbf{e}_i$

Definition: Height Ridge Lines in 2D

Definition from [Eberly 1996]:

- Assume a 2D scalar field with gradient **g** and Hessian **H** ($\mathbf{e}_i$ / λ_i s.t. $\lambda_1 \geq \lambda_2$)
- At a **local maximum**, $|\mathbf{g}|=0$, $\lambda_i<0$
  - Only true in isolated points
- To get **ridge lines,** only ask for vanishing $|\mathbf{g}|$ in direction of strongest concavity

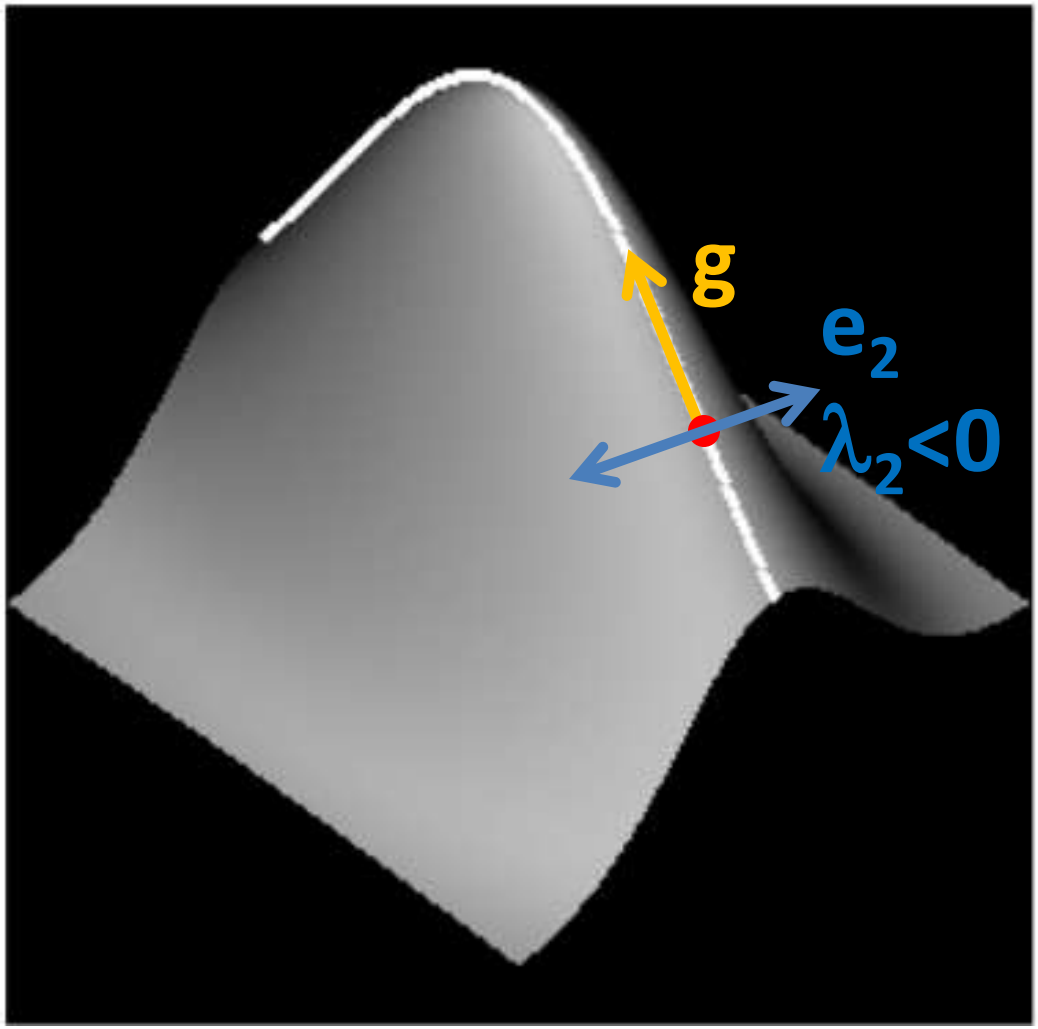
$$\mathbf{g} \cdot \mathbf{e}_2 = 0 \wedge \lambda_2 < 0$$

- **Valleys** are defined in complete analogy

$$\mathbf{g} \cdot \mathbf{e}_1 = 0 \wedge \lambda_1 > 0$$

- Ridges + Valleys are collectively called **Creases**

Illustration: Height Ridge Lines in 2D



- On Ridge Line

Height Ridge Lines and Surfaces in 3D

- Assume a 3D scalar field with gradient **g** and Hessian **H** ($\mathbf{e}_i$ / λ_i s.t. $\lambda_1 \geq \lambda_2 \geq \lambda_3$)
- At a **local maximum** (point), $|\mathbf{g}|=0$, $\lambda_i<0$
- Asking for vanishing $|\mathbf{g}|$ in direction of strongest concavity leads to a **ridge surface**

$$\mathbf{g}\cdot\mathbf{e}_3=0 \wedge \lambda_3<0$$

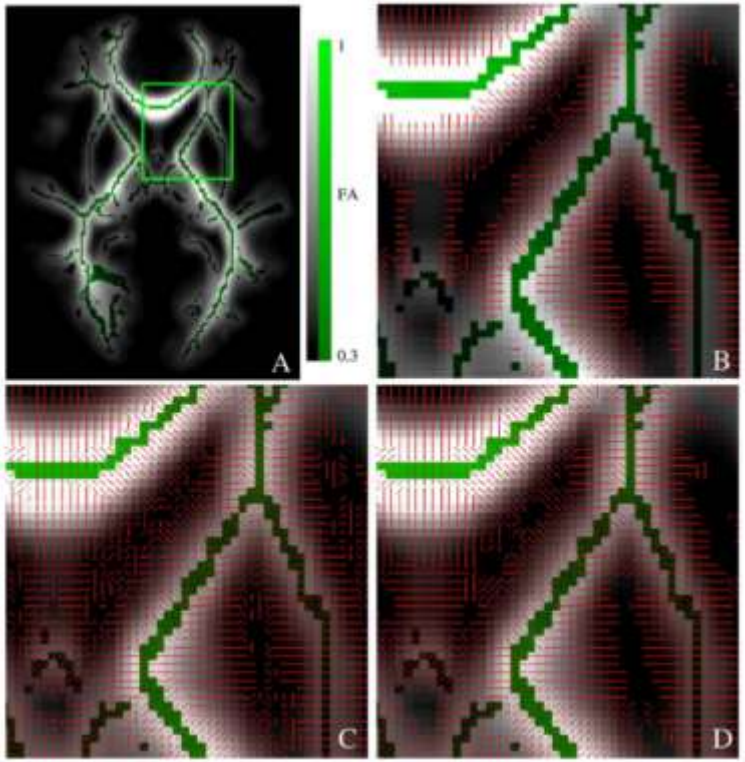
- To get a **ridge line,** ask for vanishing $|\mathbf{g}|$ in *plane* of strongest concavity

$$\mathbf{g}\cdot\mathbf{e}_3=0 \wedge \mathbf{g}\cdot\mathbf{e}_2=0 \wedge \lambda_3<0 \wedge \lambda_2<0$$

- **Again, valleys** are defined in complete analogy

Projection To The Ridge: Practice

- Build vector field for projection from



1. Gradient direction **g** (where strong enough)

2. Direction $\mathbf{e}_3$ of strongest concavity (elsewhere)

3. Slight smoothing

- Justification: Slight inaccuracies *along* the skeleton can be tolerated

Image from [Smith et al. 2006]

Projection To The Ridge: Sample Results

- Image shows the same mean skeleton (yellow) on six different individuals
  - Arrows point to voxel *from which value was taken*

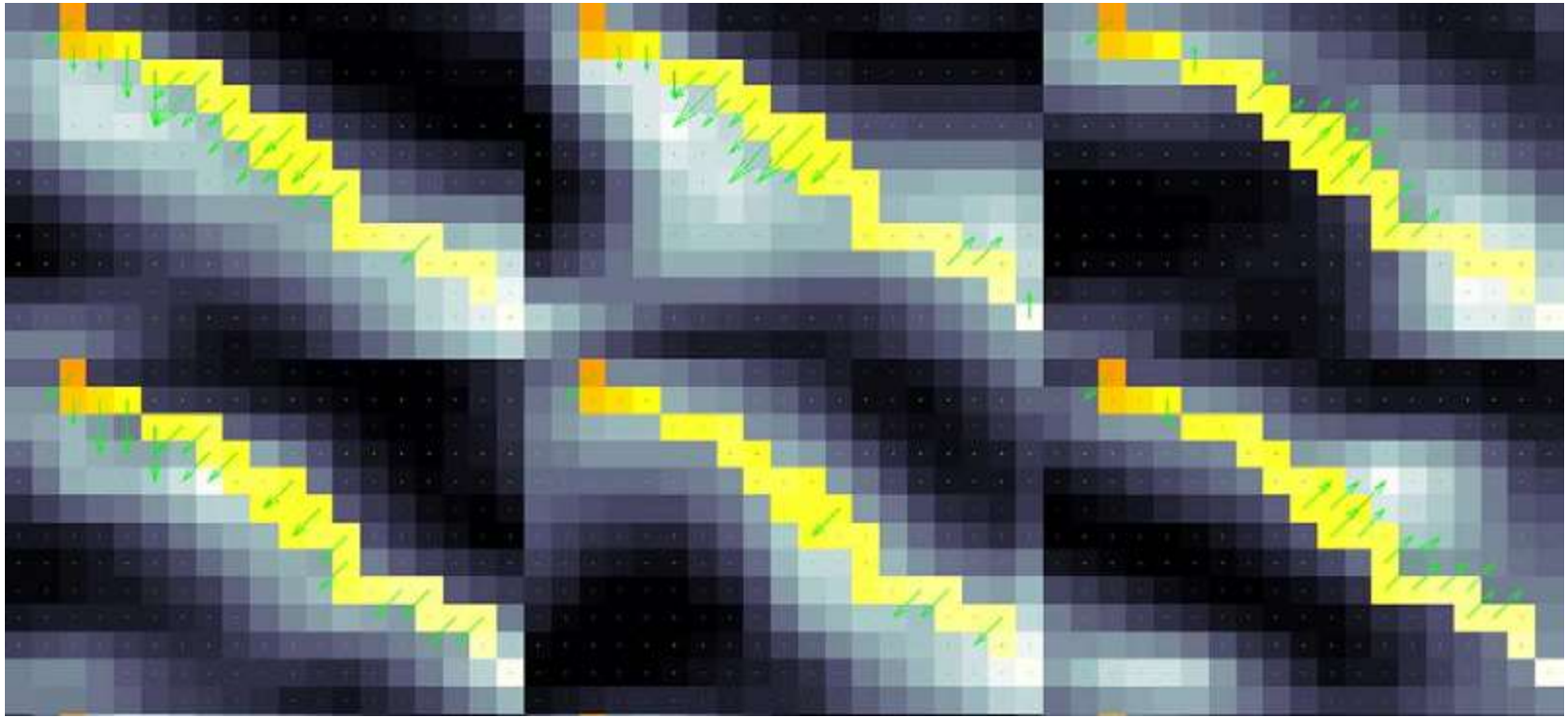
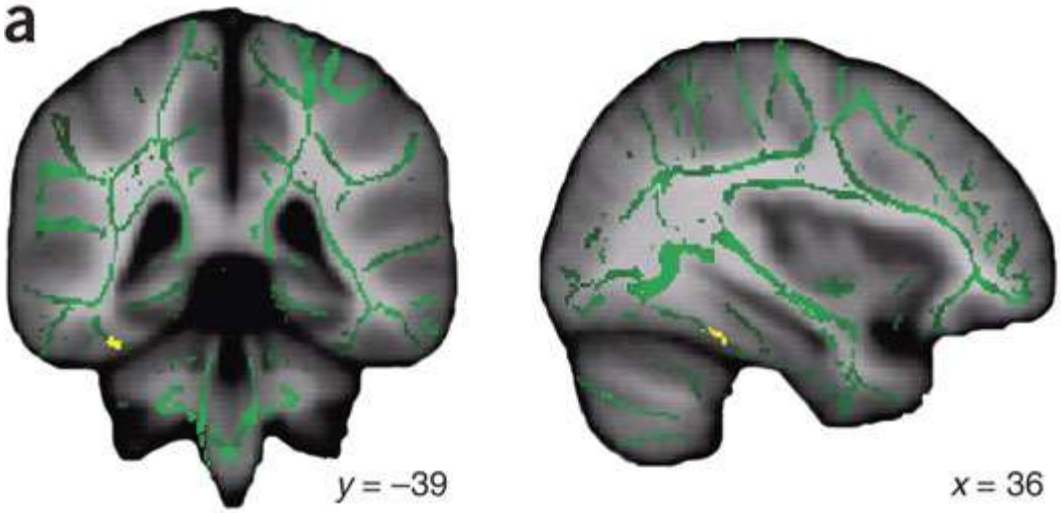
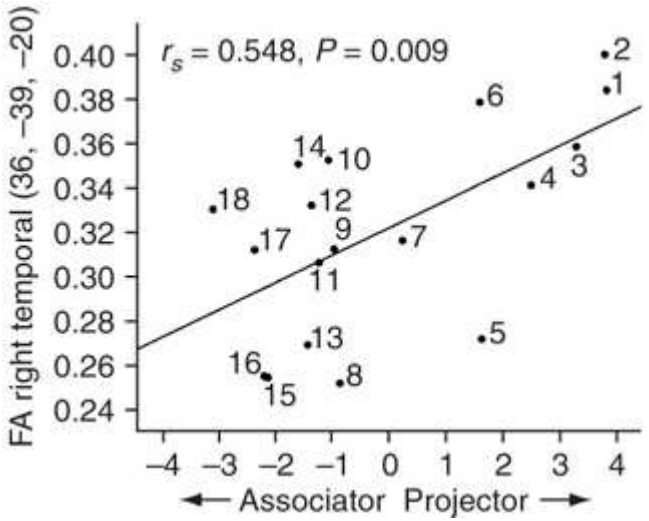


Image from [Smith et al. 2006]

TBSS: Synesthesia

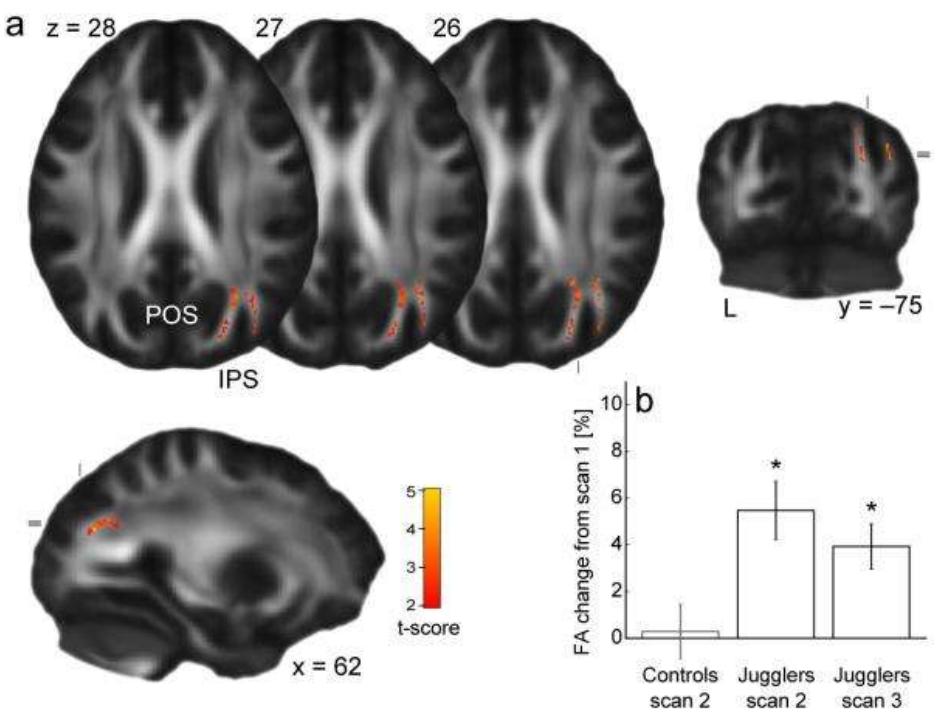
- [Rouw/Scholte 2007] showed that **grapheme-color synesthesia** is associated with increased FA in several white matter regions
  - 18 synesthetes vs. 18 normal controls
  - FA in fusiform gyrus correlated with "strength" of synesthetic experience





TBSS: A Juggling Study

- [Scholz et al. 2009] showed that **learning how to juggle** not only leads to observable changes in gray matter density, but also to increased FA in adjacent white matter
  - 48 subjects assigned to equally sized training and control groups
  - Longitudinal comparison (across time)



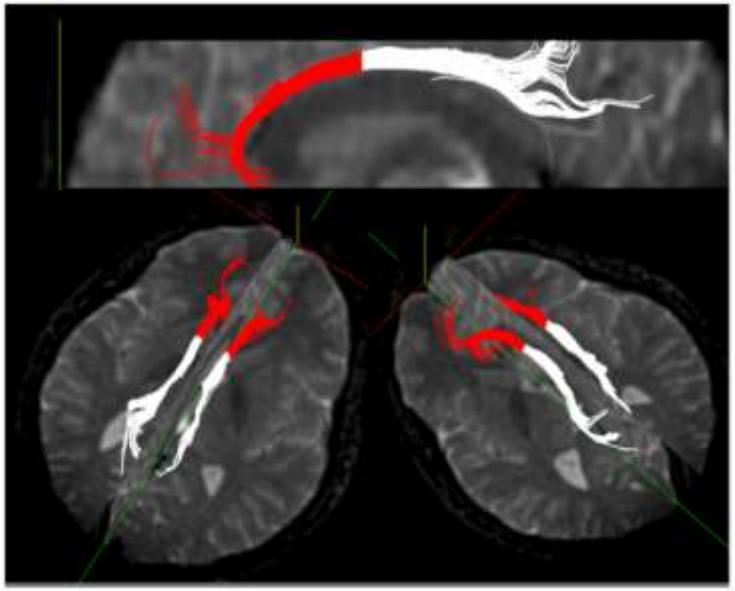
Summary: Skeletonization

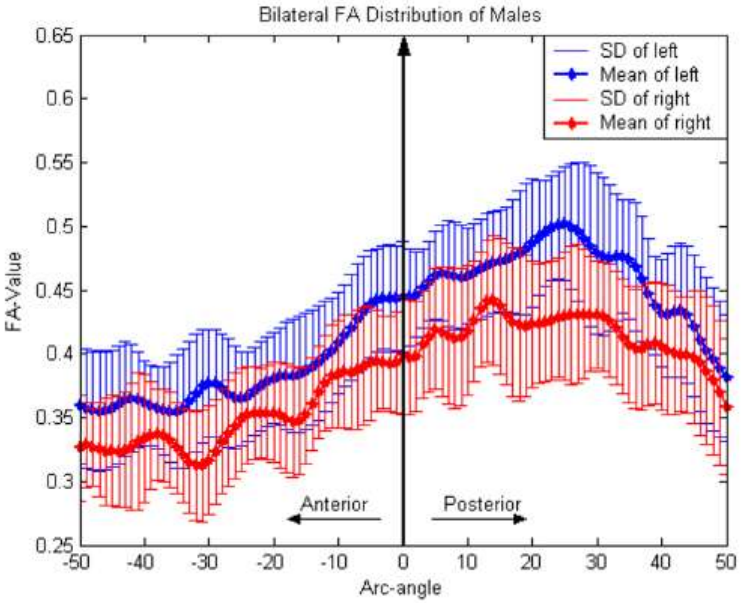
- **Problem:** In statistical analysis of FA maps, we would like to compensate incomplete alignment without smoothing
- **Tract-Based Spatial Statistics** is the standard approach to this problem
  - Restricts statistical analysis to a “skeleton”
  - Projection to that skeleton compensates for initial misalignment

10.3 Fiber Tracking

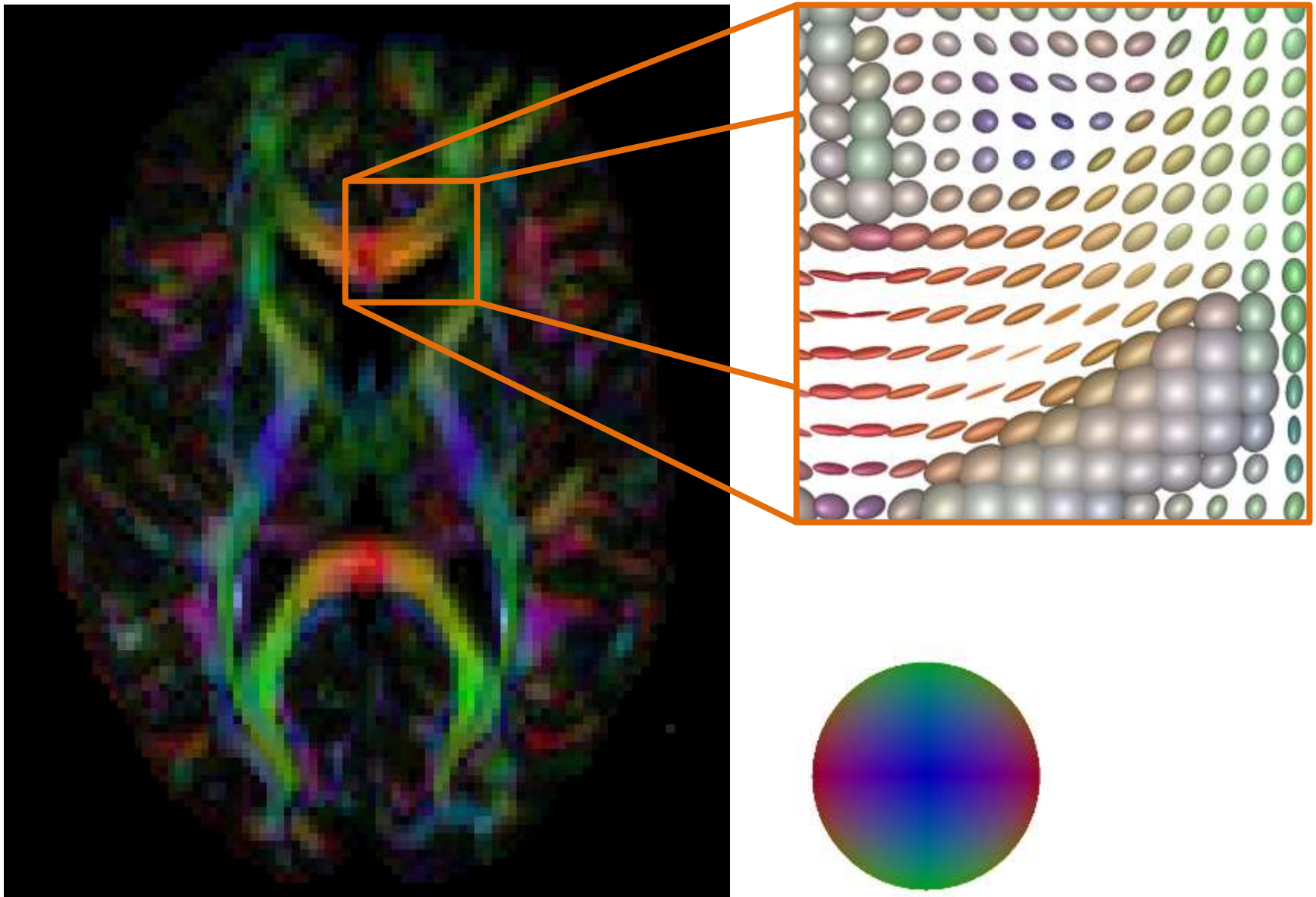
Motivation: Fiber Tracking

- **Fiber tracking / tractography** uses inferred local fiber directions to reconstruct the trajectories of major nerve fiber pathways
  - Find out in which bundle a difference was localized
  - Perform statistical analysis along a bundle
  - Study fiber geometry itself (e.g., particularly of interest when planning surgery)



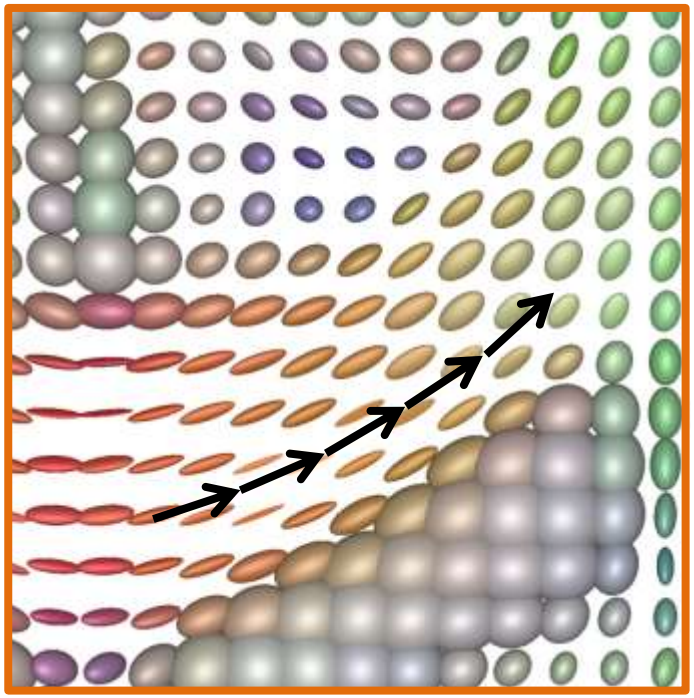


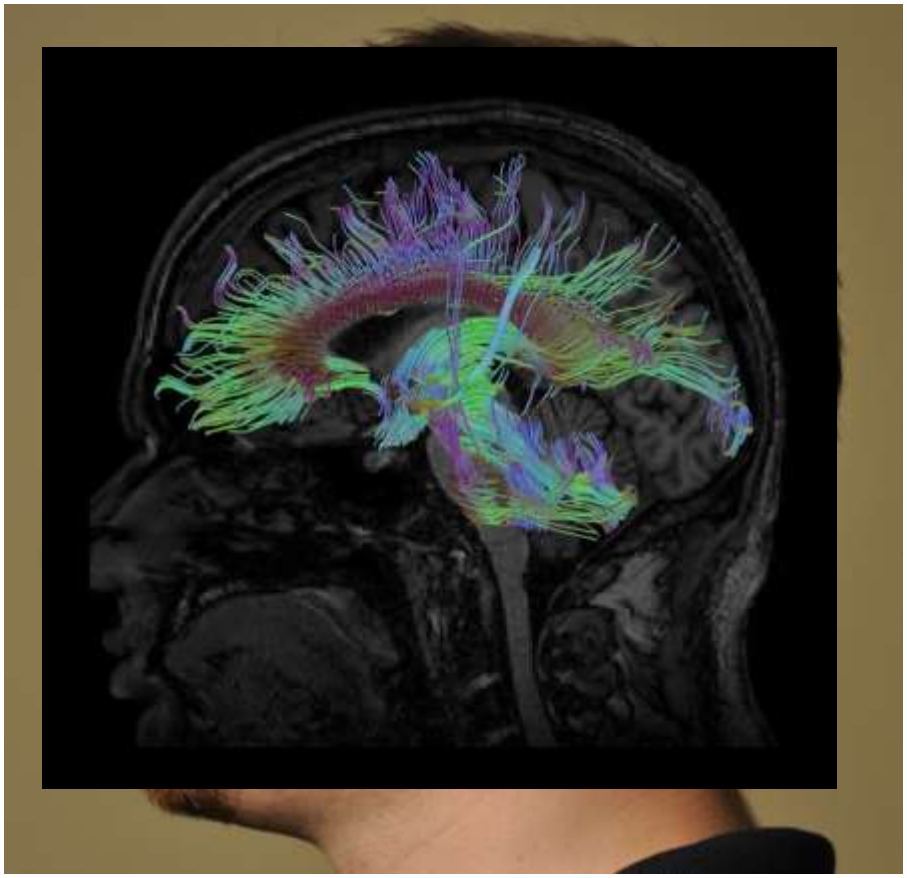
Images from [Gong 2005]



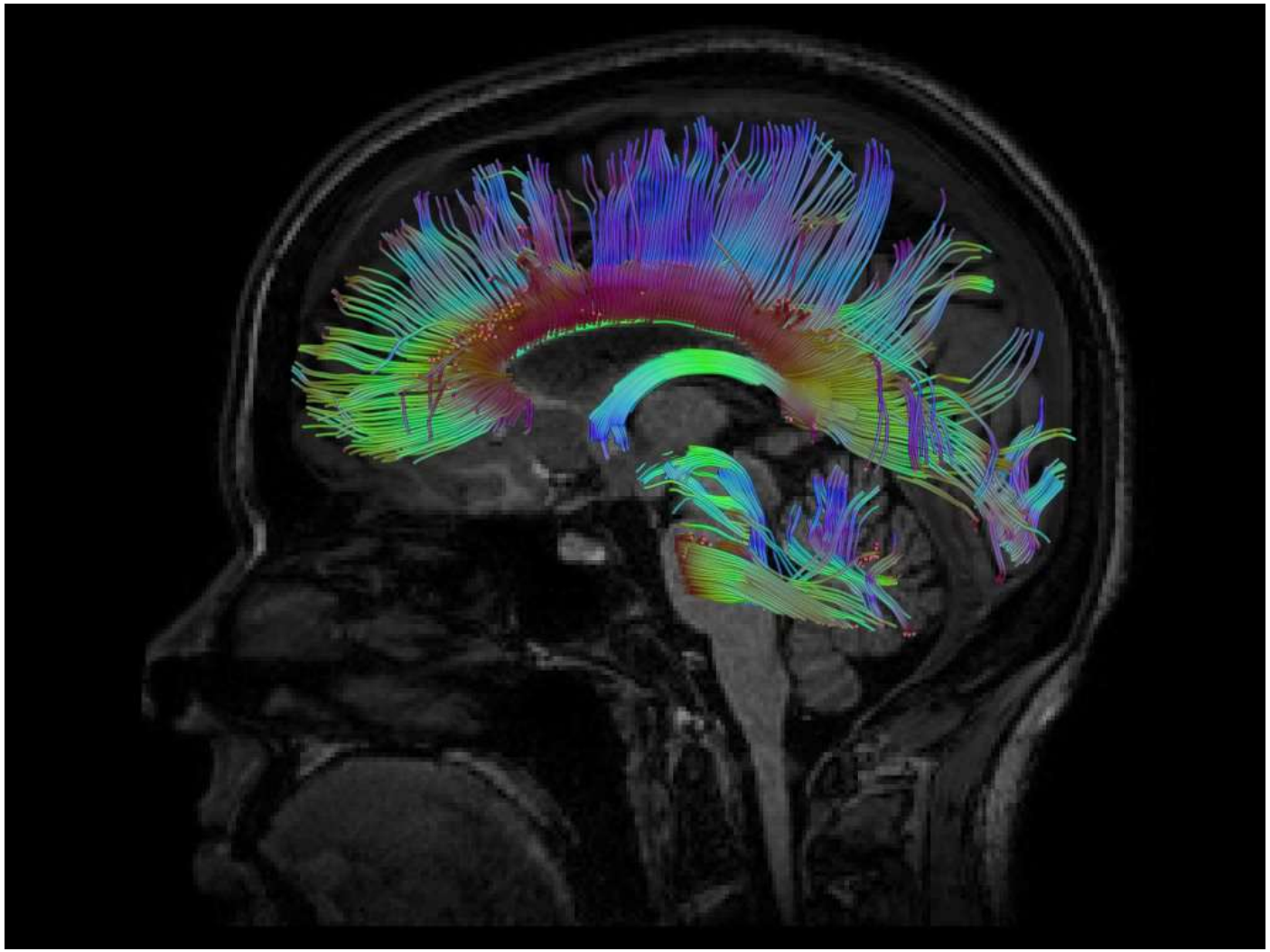
Deterministic Fiber Tracking

- In **deterministic tractography,** we proceed along a fixed tangent direction in each step





Sample Result: Deterministic Tracking



Fiber Tracking as Solution of an ODE

- Deterministic Tractography can be viewed as solving the ordinary differential equation

$$\dot{\mathbf{x}}(t) = \mathbf{v}\big(\mathbf{x}(t)\big)$$

  - Vector field **v** derived from dMRI model
  - *Example:* Major eigenvector of diffusion tensor
  - Choose sign of **v** so that we are tracking forwards
- **Basser et al. [2000]:**
  - **Euler integration** with stepsize s: $\mathbf{x}_{i+1} = \mathbf{x}_i + s\ \mathbf{v}(\mathbf{x}_i)$
  - More exact: Higher-order schemes **(Runge-Kutta)**
  - Stop on **high curvature** or **low anisotropy**
  - Use interpolation to obtain **continuous** tensor field
  - At each step, **interpolate tensor** and compute its principal eigenvector

FACT

Fiber Assignment by Continuous Tracking (FACT)

[Mori et al. 1999]

- Follow principal eigenvector until voxel is left
- Stop when direction would change abruptly

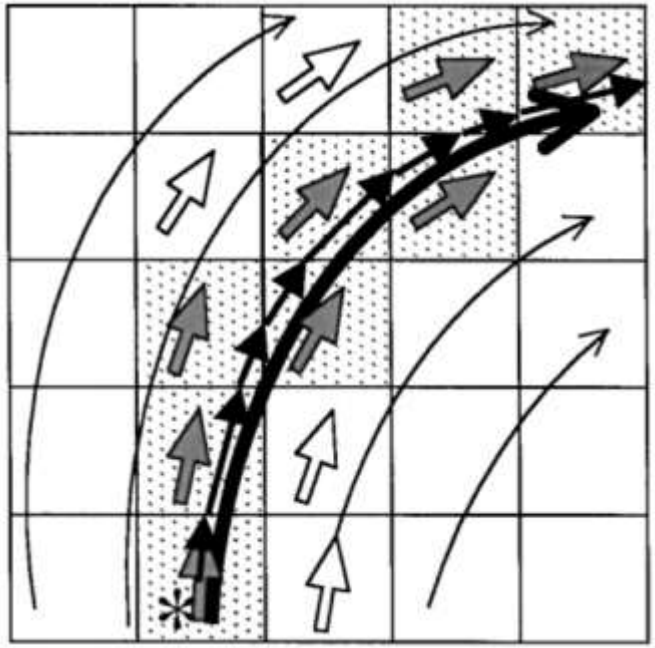
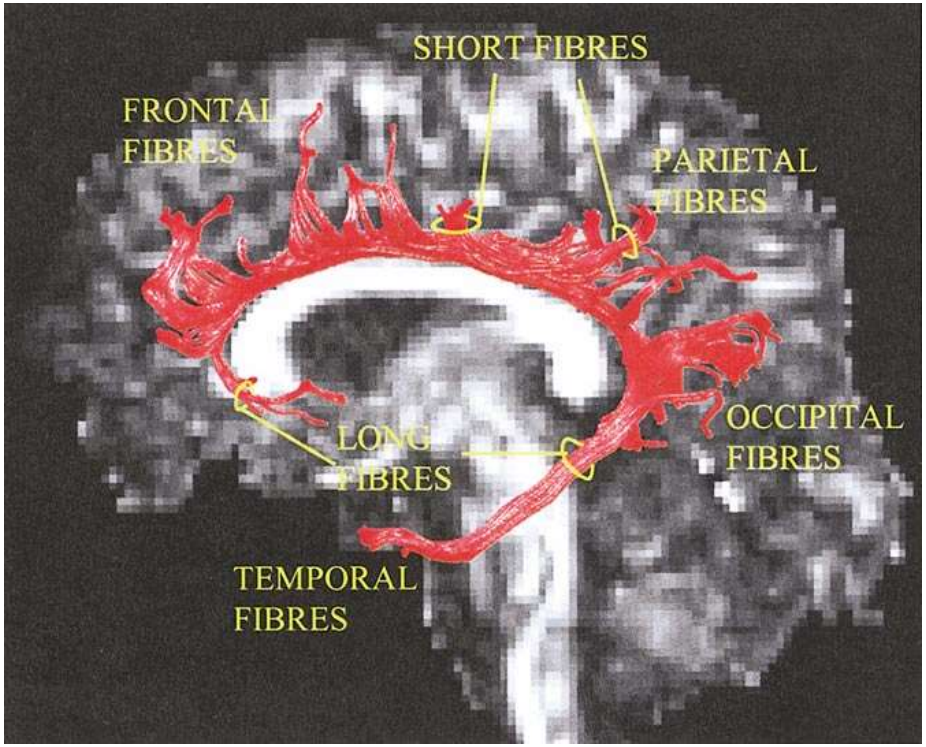


Image taken from Mori et al. [1999]

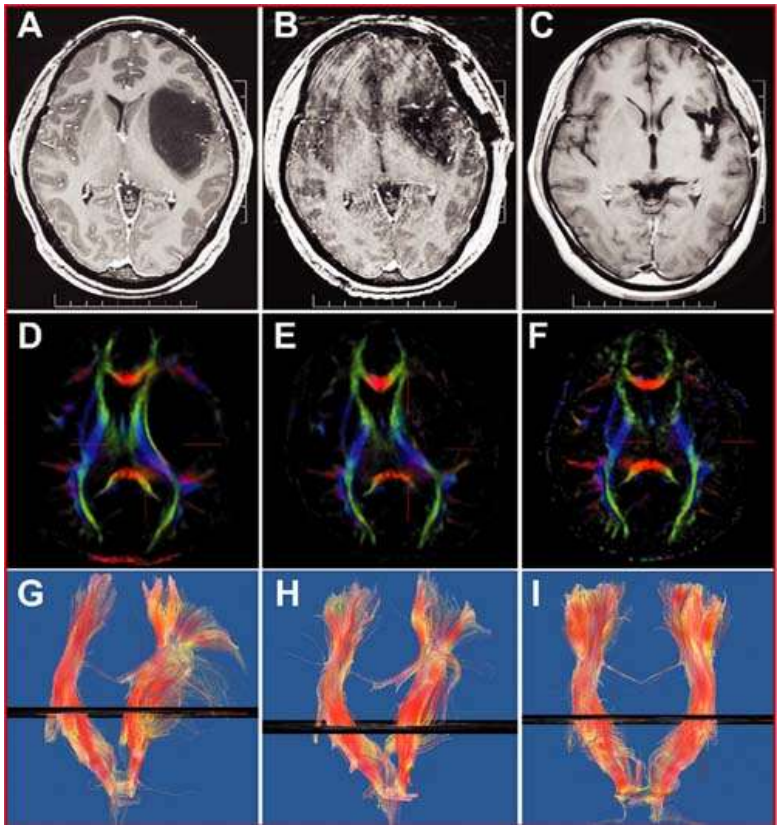
Success: White Matter Atlases

- **[Catani et al. 2002]** Deterministic tractography reliably reconstructs many previously known large-scale white matter tracts
  - Previously not possible to study them *in vivo*



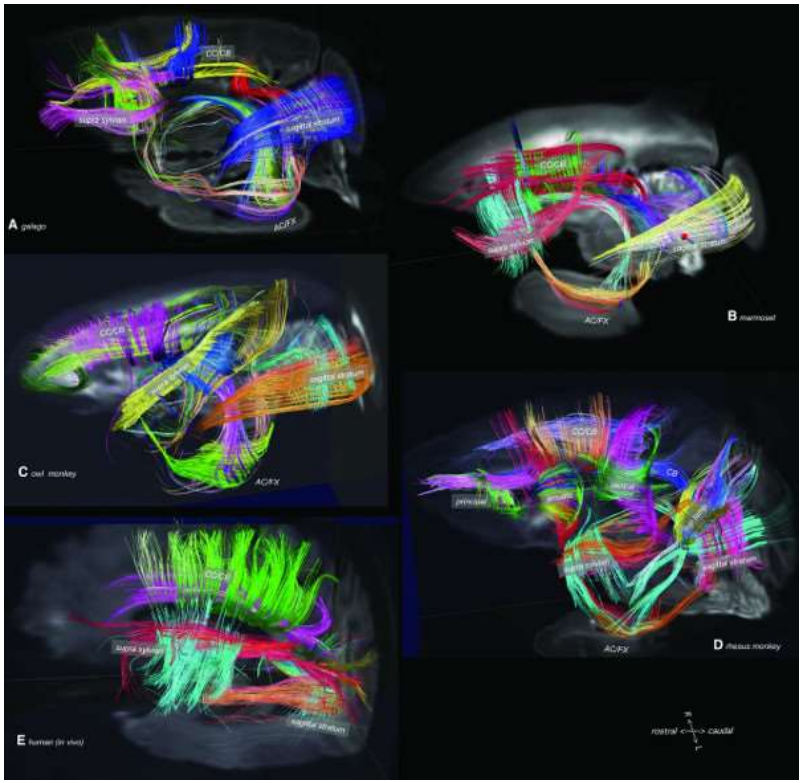
Success: Surgical Planning

- **[Wu et al. 2007]** randomly either used (n=118) or did not use (n=120) dMRI for planning brain tumor surgery. They report that dMRI led to
  - Greater chance of complete resection
  - Reduced chance of postoperative motor deterioration
  - Longer survival
- **But:** Due to software constraints, mostly based on FA maps, fiber tracking only used in 23 cases



Success: White Matter Architecture

- **[Wedeen et al. 2012]** used fiber tracking for a comparative study of white matter geometry in humans and several non-human primates
  - 3D grid of parallel sheets of near-orthogonally interwoven paths
  - Strong cross-species homologies
  - Postulate three principal axes of development



Pros and Cons of Deterministic Tracking

Advantages of Deterministic Tractography:

- Relatively simple and fast
- Successfully reconstructs many known bundles
- Leads to crisp visualizations

Limitations of Deterministic Tractography:

- No way to distinguish between highly reproducible streamlines and noise-induced false positives
- Streamlines **do not** have an anatomical counterpart
  - Individual axons are much smaller
  - Fiber bundles are not point-to-point connections

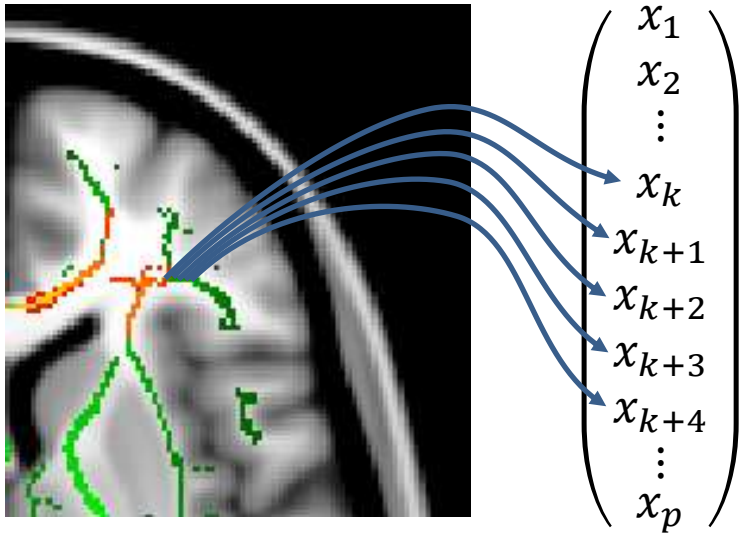
Summary: Diffusion MRI

- During its **spontaneous heat motion,** water interacts with tissue at (sub-)cellular level
- **Diffusion MRI** probes this restricted diffusion behavior to draw conclusions about structures that are far below image resolution
  - **Scalar indices** of tissue microstructure (e.g., Fractional Anisotropy)
    - Statistical analysis often involves skeletonization
  - **Tractography:** Reconstruction of white matter pathways
    - Multi-fiber models to track through crossings
  - Additional contrast even within **gray matter**

10.4 BundleMAP

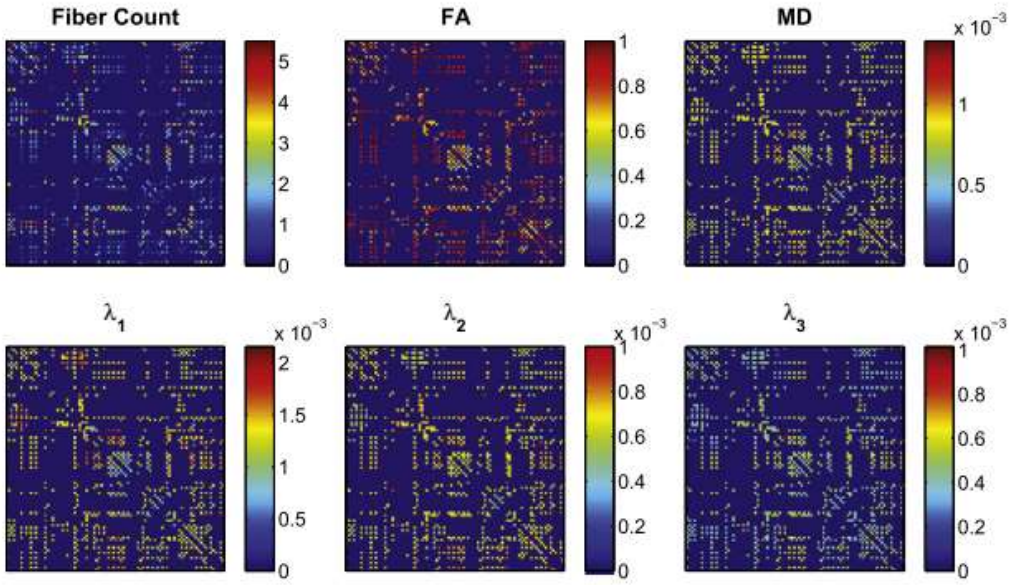
One Feature Per Voxel?

- **1st Traditional Approach:**
  One feature per voxel and metric
  - **Consequence:**
    „small n, large p“
    - ≈60 subjects
    - ≈120.000 features,
      most of which are irrelevant
    - Feature selection / feature weighting required



One Feature Per Tract?

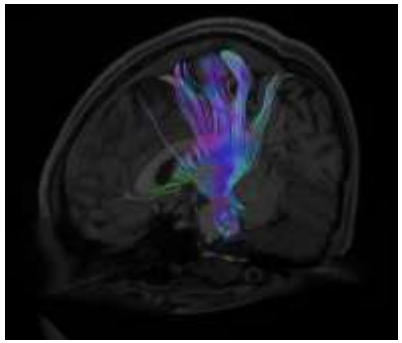
- **2nd Traditional Approach:**
  Features from connectivity matrices
  - Consequence:
    - One feature per tract and metric
    - Might water down power if only some part of a (long) tract is affected

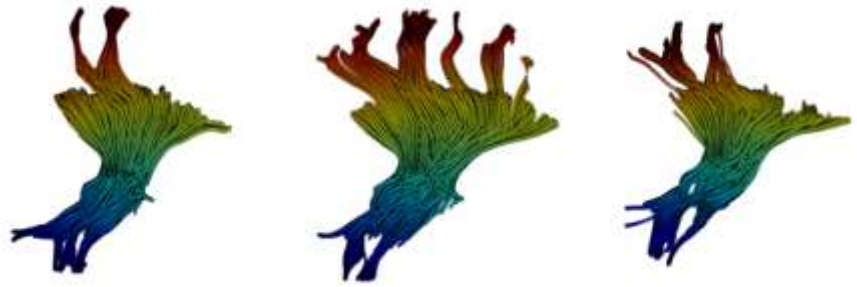


[Wee et al. 2012]

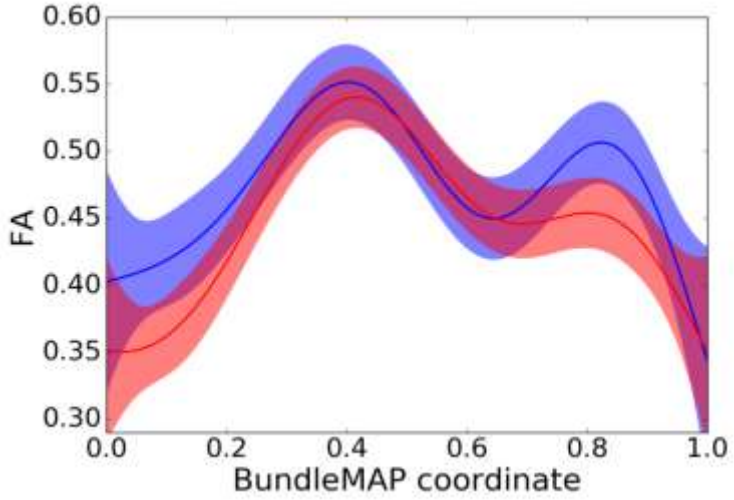
Along-the-tract Feature Extraction

- **BundleMAP:** Automatic anatomical correspondences between subjects



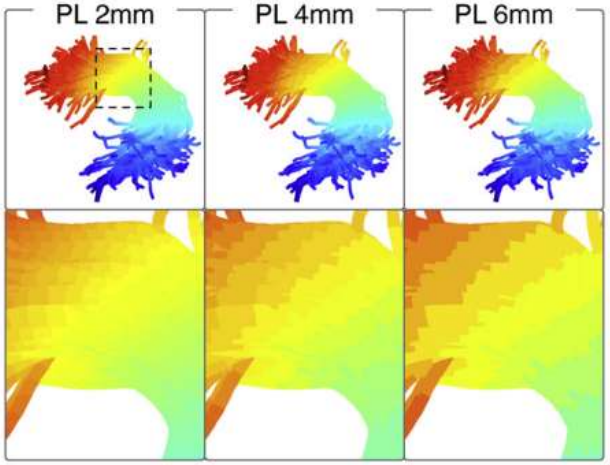


- Simplifies visual and statistical comparison:
  - Patients (NPSLE)
  - Healthy Controls

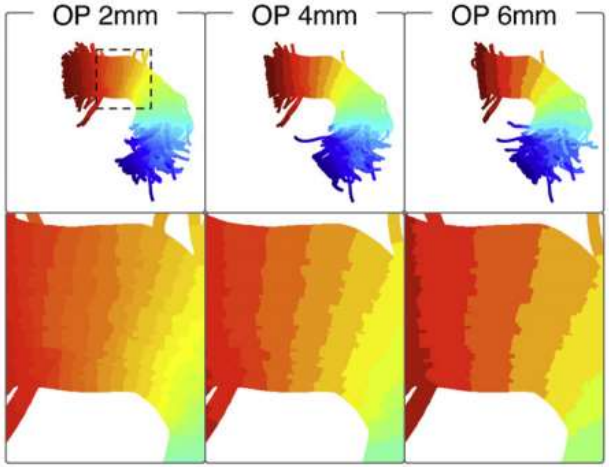


Prior Tools for Along-the-tract Analysis

- Manually specified **cutting plane** and arc-length
  - [Corouge et al. 2006]
  - [Zhu et al. 2010]

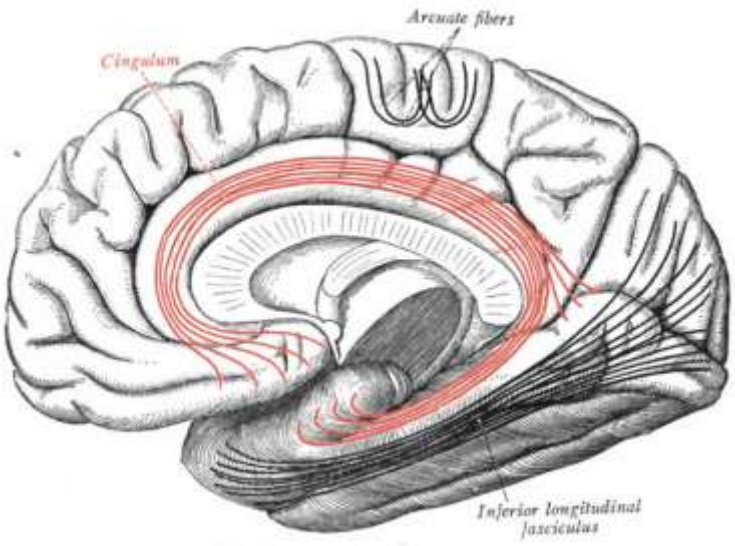


- Selection of **prototype fiber** and optimal matching of all others (Hungarian algorithm)
  - [O'Donnell et al. 2009]

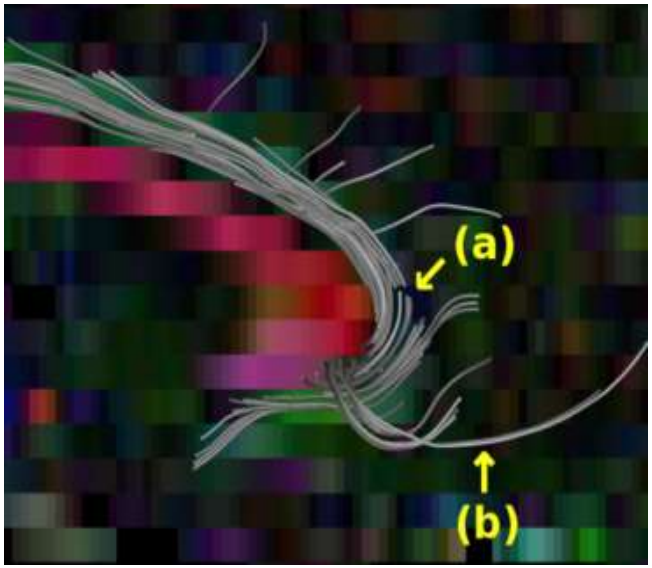


A New Perspective

- Map curves to an **idealized fiber bundle core**
  - Latent 1D manifold, connecting two brain regions
  - Can be established using manifold learning
  - We make use of modified ISOMAP, thus the name BundleMAP



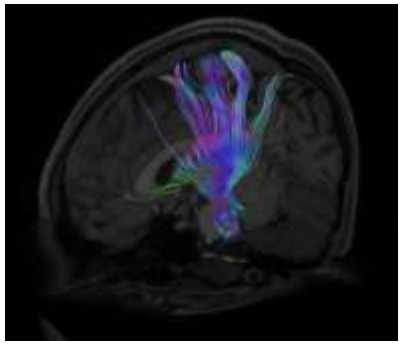
Sobotta‘s Anatomy / wikimedia



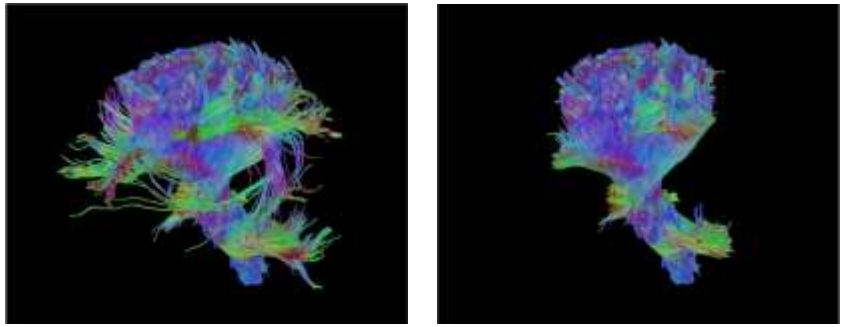
Tracking on clinical data

BundleMAP Pipeline

1. Tractography

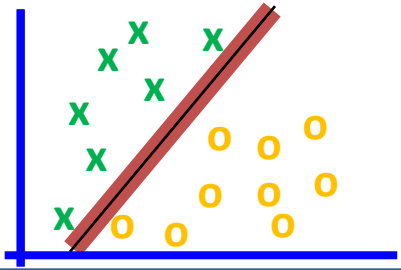


2. Outlier Removal



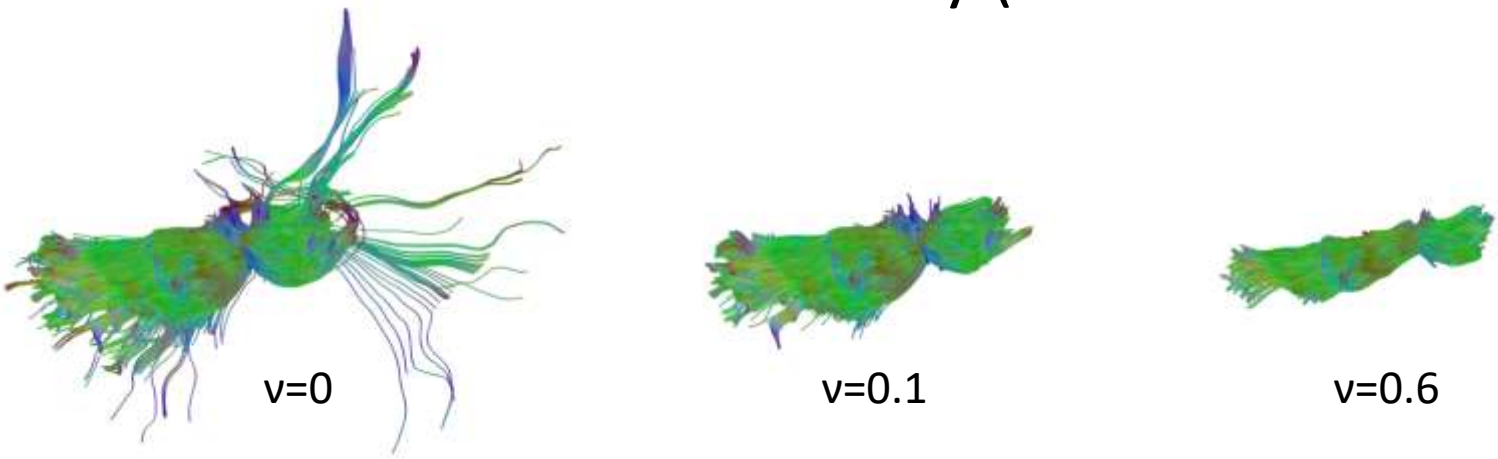
3. Manifold Learning

4. Adaptive Binning, SVM



Tractography and Outlier Removal

- **Tractography** fully automated
  - Seeds set by registration to a template
  - Rough pre-alignment of fibers using same transform
- **Outlier removal** using one-class SVM
  - Represents each curve using mean+covariance [Brun et al. 2004]
  - Choice of ν validated manually (ν=0.1 almost always)



Manifold Learning

- Modifications to ISOMAP:
  - Apply a custom distance function
    - arc-length along the bundle
    - Projection on tangent otherwise

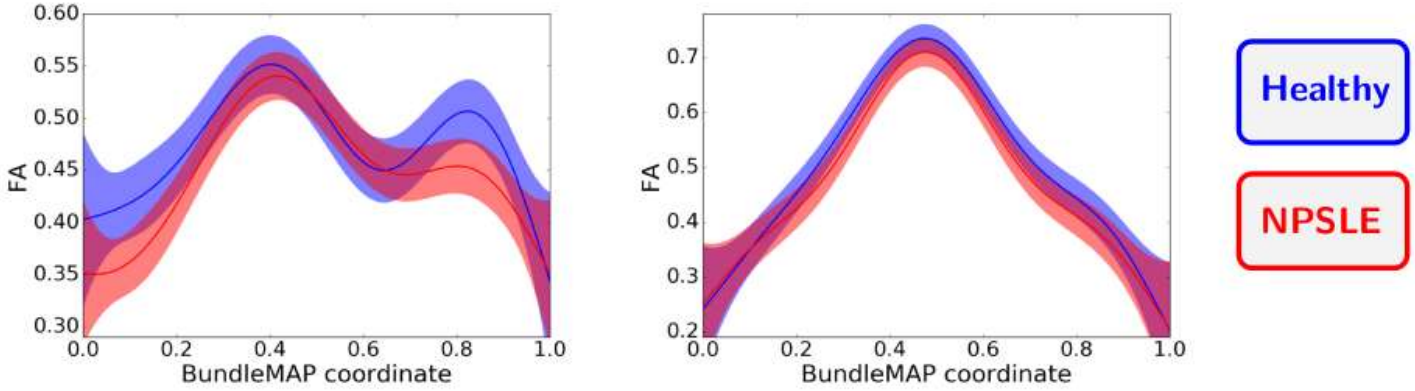
$$d(\mathbf{v}_1, \mathbf{v}_2) := \frac{1}{2} \sum_{i=1}^{2} \left| \frac{\mathbf{t}_i \cdot (\mathbf{v}_2 - \mathbf{v}_1)}{\|\mathbf{t}_i\|} \right|$$

  - Normalize ISOMAP result to [0,1]
  - Fix sign so that coordinates run left-to-right, back-to-front, bottom-to-top
  - Only apply ISOMAP to a subset of curves (cluster centers from k-means)
  - Results interpolated to all curves

Results from Manifold Learning

- Color-coded correspondences:

- FA as a function of position along the tract:



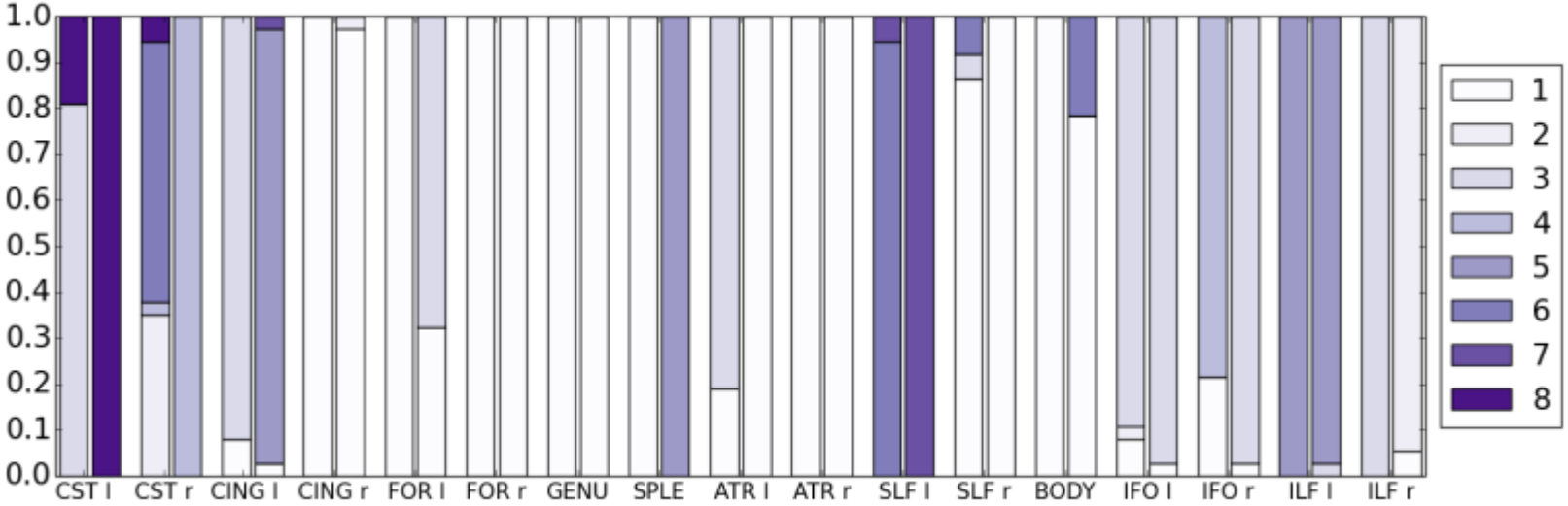
(a) Left IFO (b) Genu of the *corpus callosum*

Adaptive Binning

- **Most suitable bin resolution** will depend on tract and disease
  - We select the binning that maximizes

$$F = \frac{\|\mathbf{x}^{(p)} - \mathbf{x}^{(a)}\|^2 + \|\mathbf{x}^{(n)} - \mathbf{x}^{(a)}\|^2}{\frac{1}{|\mathcal{P}|-1}\sum_{\mathbf{x}_i \in \mathcal{P}} \|\mathbf{x}_i - \mathbf{x}^{(p)}\|^2 + \frac{1}{|\mathcal{N}|-1}\sum_{\mathbf{x}_i \in \mathcal{N}} \|\mathbf{x}_i - \mathbf{x}^{(n)}\|^2}$$

  - Choice quite stable in cross-validation and clearly dependent on bundle:



Example Application

- **56 subjects** from study on SLE (lupus)
  - 19 with neuropsychiatric symptoms (NPSLE)
  - 19 without NP symptoms (non-NPSLE)
  - 18 healthy controls
- 15 gradient directions at b=800 s/mm²
- **17 major bundles** throughout the whole brain, no specific prior hypothesis

Benefit for Classification and Regression

- **Improved Classification Accuracy:**

	HC vs. non-NPSLE	HC vs. NPSLE
TBSS + F-Scores	70%	70%
Tract-Based	54%	70%
BundleMAP	**73%**	**76%**

- **Regression:** Prediction of SLEDAI-Scores (range 0-24)
  - Baseline (blind): 5.43±3.49
  - Tract-based: 5.12±3.83
  - BundleMAP: **4.65±3.86**

Ranking of Relevant Bundles

- **Anatomical Interpretability:**

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
HC vs. non-NPSLE	FOR r	FOR l	ATR r	BODY	CING l
HC vs. NPSLE	ATR r	IFO l	GENU	ATR l	ILF r
SLEDAI	ILF l	CST l	SLF l	IFO r	ATR l

  - Good agreement with results from different method applied to independent data [Emmer et al. 2010]
  - Fornix very plausible, but was not detected in TBSS analysis of same data
    - Our method was likely more successful to bring this thin structure into alignment [Bach et al. 2014]

Comparison to Arc-Length

- **Simple and widely used** alternative for parametrization:
  - **Arc-length distance** from a manually defined origin
  - Requires extrapolation
  - Led to poorer accuracy (68% in both cases)
  - Manifold learning adds robustness

Arc-Length Result

BundleMAP Result

Conclusion

- Propose to treat **joint tract parametrization** as a **manifold learning problem**
  - Fully automated and robust result
  - Simple implementation (based on ML packages)
- **Integrated framework** for supervised learning
  - Includes outlier removal and adaptive binning
  - Relatively good predictive power
  - Can also be used for statistics and visualization
  - Highlights most relevant bundles

Remainder of This Lecture

- Second project will be discussed on **Tue Feb 14 at 9:30 at Marschallsaal, B-IT**
- **Tue Feb 21 at 10:00 at LBH I.80:** Question & Answer Session
- **Fri Feb 24 at 9:30:** Written exam in B-IT Lecture Hall
  - You can bring one A4 sheet of paper with hand-written notes
- **Tue Mar 14 at 9:30:** Re-exam in B-IT Lecture Hall

Further Reading

- D.K. Jones (editor): *Diffusion MRI: Theory, Methods, and Applications.* Oxford University Press, 2011
- Tournier, Mori, Leemans: *Diffusion Tensor Imaging and Beyond.* Magnetic Resonance in Medicine 65(6):1532-1556, 2011