# Fast diffusion imaging using compressed sensing in q-space

Yogesh Rathi
Assistant Professor
Brigham and Women's Hospital,
Harvard Medical School, Boston

Outline

- Why fast diffusion imaging ?
- Approaches to fast diffusion imaging
  - Fast acquisition sequences
  - Smart signal reconstruction
- Compressed sensing
- Spherical Ridgelets and extensions
- Experiments and Results

Background – diffusion MRI
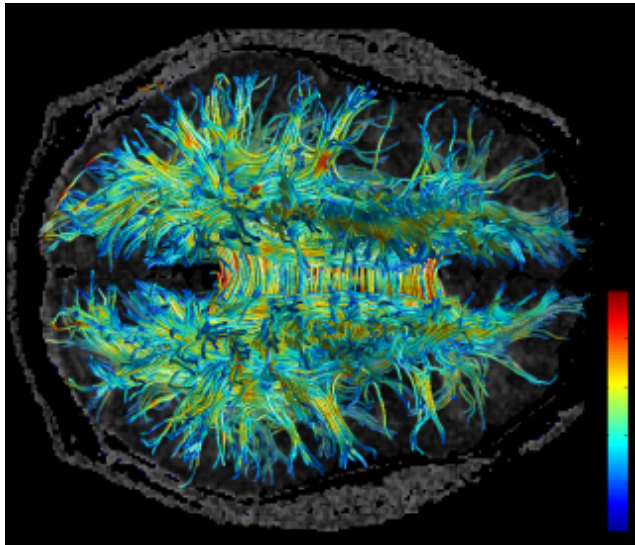
Diffusion MRI (dMRI) allows non-invasive investigation of neural architecture of the brain. It is one of the most widely used mechanisms to study several brain disorders.

dMRI and mental disorders

- dMRI used to identify location of stroke
- Used in characterizing edema (Pasternak, 2009)
- High b-value data was very sensitive to disease load in MS patients in characterizing normal appearing white matter (Cohen 2002)
- Q-space sampling allows to characterize the fast and slow diffusing components which may belong to different tissue regions (AxCaliber – Assaf et al)
- Diffusion Kurtosis Imaging (DKI) has been shown to be more sensitive to tissue changes (neurodevelopmental and in mild TBI)

Background - dMRI
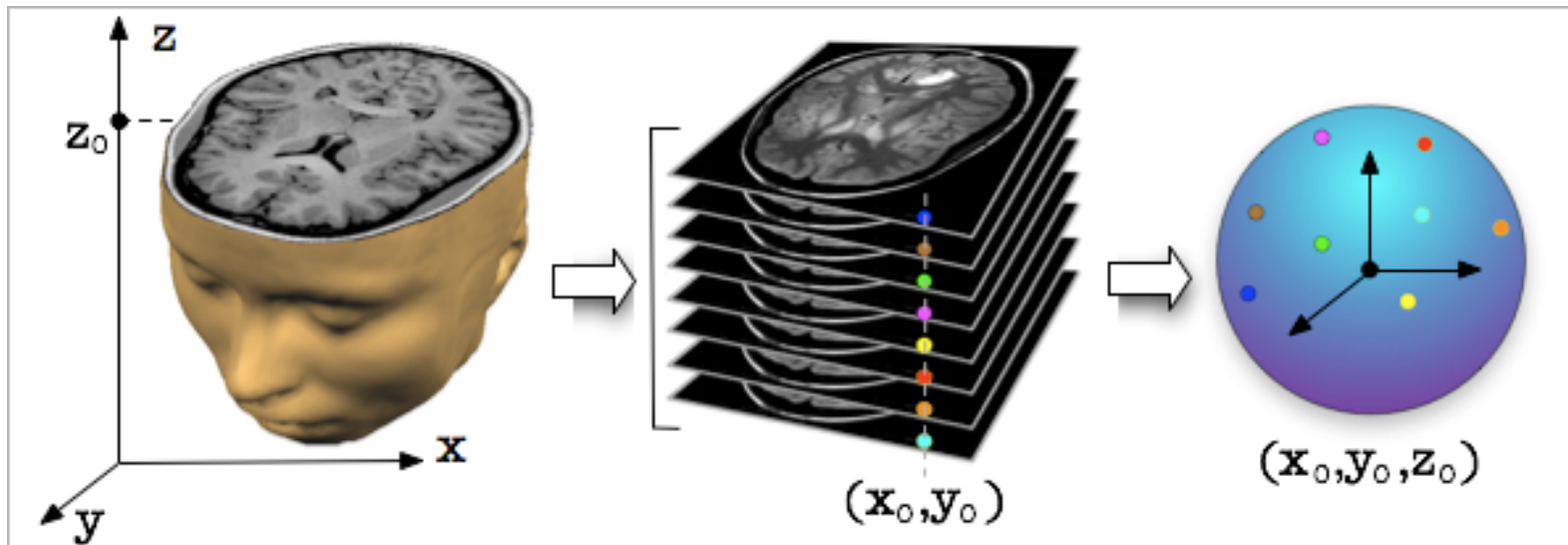
- Typically, diffusion tensor imaging is used in clinical settings (requires only 7 gradient directions).
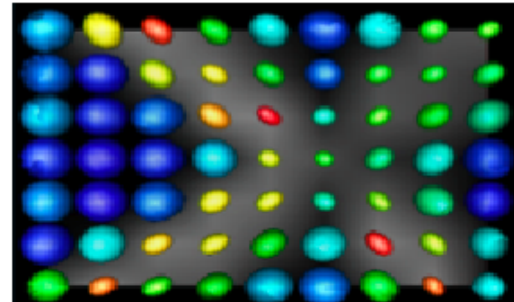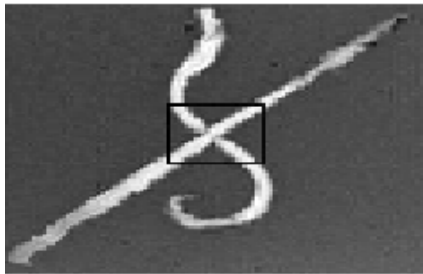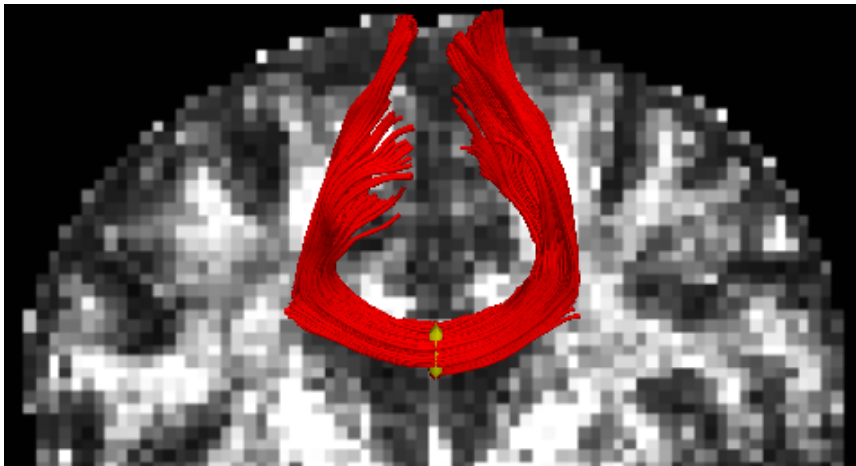
Diffusion Tensor Imaging
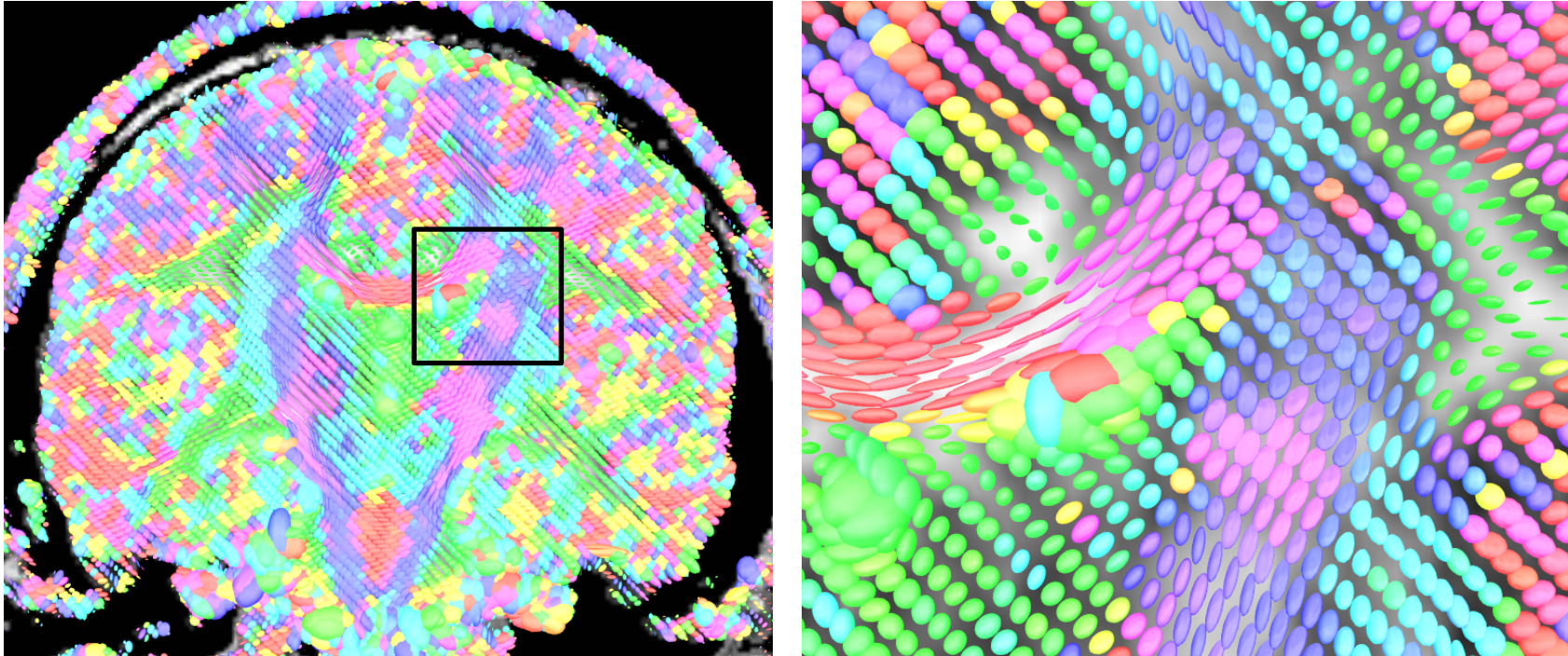
- At each location, the diffusion behavior of water is modeled as an ellipsoid.
- In medical imaging this ellipsoid is called a diffusion tensor.
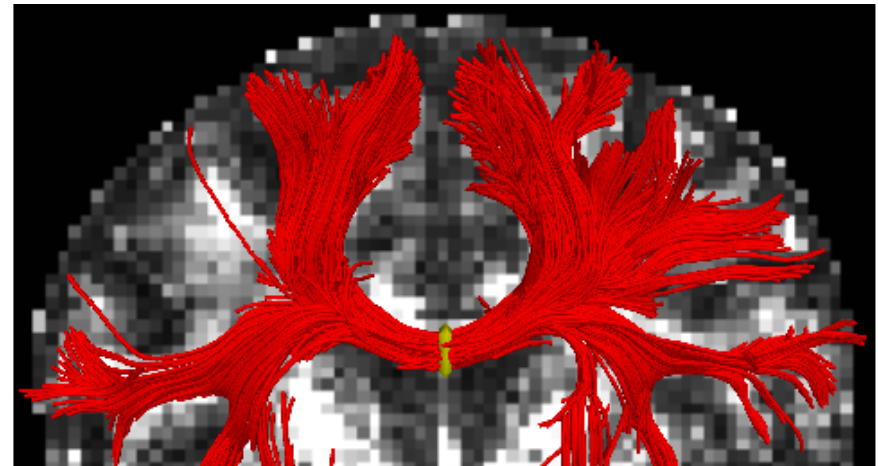
Advantages of HARDI

Aim of Tractography : In-vivo tracing of neural pathways of the brain.

[Campbell et al.]

single-tensor

filtered two-tensor

(Rathi et al.)

Advanced- dMRI

To address this problem, Tuch et al (2004), proposed High Angular Resolution Diffusion Imaging (HARDI), which involves acquiring several gradient directions uniformly spread over a sphere.

👍 With HARDI, multiple fiber crossings can be detected.

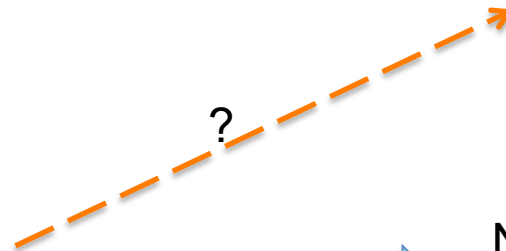👎 Acquisition time (10-20 minutes) increases significantly, since many measurements are required.
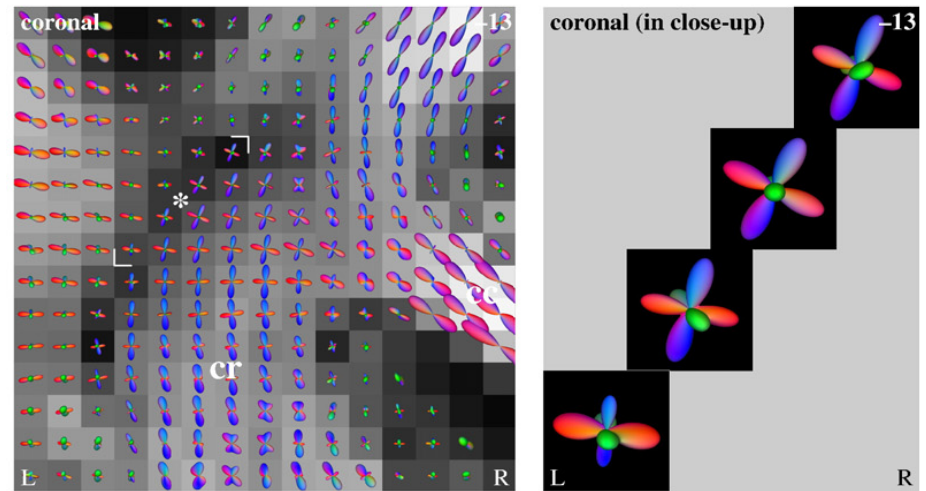
Picture from Kaden et al, Neuroimage, 2008

Time Considerations
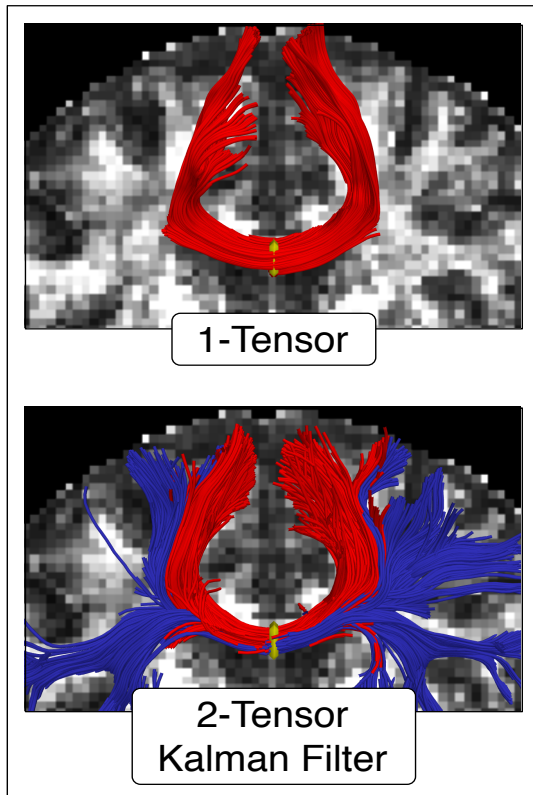
Need at-least 6 gradient directions and at-most 5 minutes of scan time.

Need more than 45 gradient directions and about 15-20 minutes of scan time.

Time considerations

- To reduce acquisition time, we can take 2 paths:
  - Faster acquisition methods
    - Multi-slice acquisition (Setsompop et al, 2010).
    - Multiplexed EPI (Feinberg et al, 2010).
  - Smarter signal reconstruction methods from fewer measurements (Michailovich and Rathi et al, 2010).

- My current work focuses on using compressive sampling and diffusion models for signal reconstruction (topic of this talk).

Background – Compressive Sampling

- Compressive Sampling or compressed sensing (CS) theory asserts that one can recover certain signals from far fewer measurements than is traditionally required as given by Nyquist criteria.

  [Candes, Romberg, Donoho, etc].

- To make this possible, CS relies on two fundamental concepts:

  – Sparsity
  – Incoherence

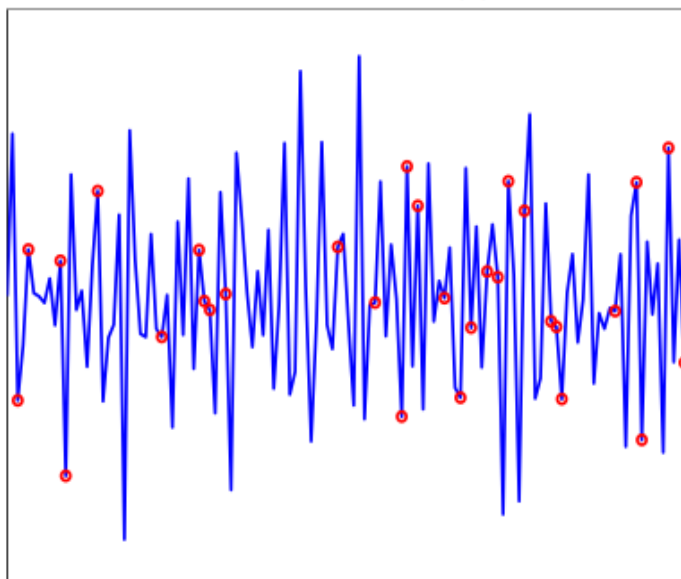# Sparsity

input signal x(t)
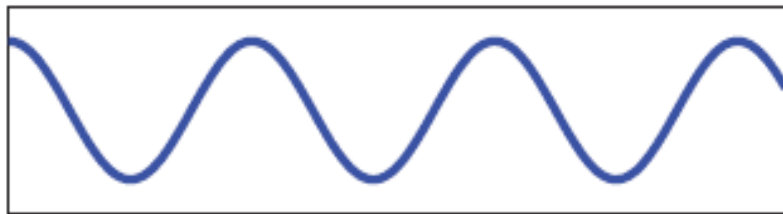
Time domain

input signal X(ω)

Frequency domain

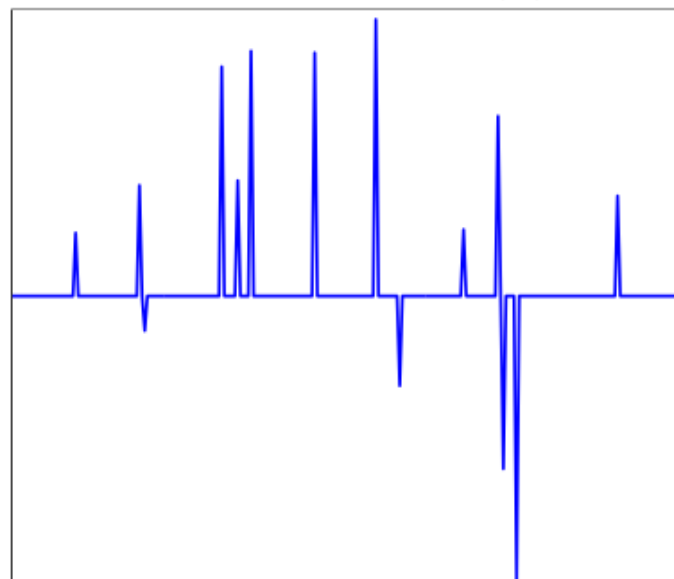CS theory <u>requires</u> that the signal of interest be sparse or compressible in some basis (domain) Ψ

Sampling Example

Time domain $f(t)$

Measure M samples

(red circles = samples)

Frequency domain $\hat{f}(\omega)$

K nonzero components

$$\#\{\omega : \hat{f}(\omega) \neq 0\} = K$$

Incoherence

- Unlike the signal of interest, the <u>sampling waveforms (basis)</u> Φ should have a very dense representation in Ψ

- Thus, if Ψ is a Fourier basis, then using a Dirac delta function as a sampling waveform would imply that its support in Ψ is very dense.

$\Longrightarrow$ Fourier and Dirac and very *Incoherent*

Incoherence

- The coherence between the sampling basis Φ and the representation basis Ψ is given by

$$\mu(\Phi, \Psi) = \sqrt{n} \max_{1 \leq k,j \leq n} | < \phi_k, \psi_j > |$$

- Higher values for $\mu(\Phi, \Psi)$ indicates more coherence: so for effective application of CS, this value should be as small as possible – ($n^{-1/2}$, 1).

Restricted Isometry (RIP), Uniform uncertainity (UUP)

- Theorem (due to Candes and Romberg, 2007) : Let $S \in R^n$ be the signal, whose representation in the basis Ψ is K-sparse (i.e., only K coefficients are non-zero):

- Then, with overwhelming probability, the number of measurements m required to exactly recover the signal is given by:

$$m \geq C.\mu^2(\Phi, \Psi).K.\log n$$
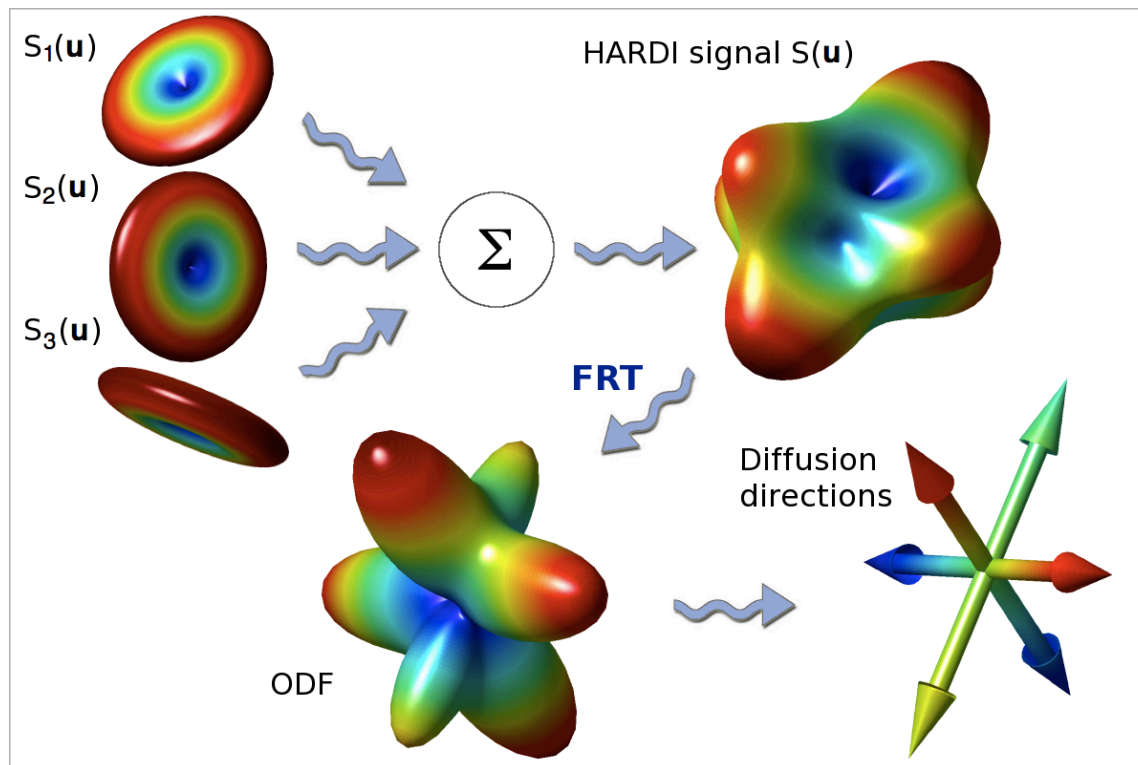
where, C is a positive constant.

Incoherence

- In the case of dMRI, we can assume that the sampling basis is fixed , i.e.

  The Dirac Delta function

- So, we need a basis

  – that provides a sparse representation of the dMRI signal and

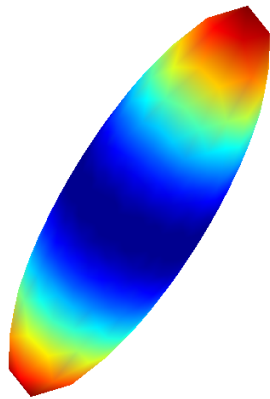 – has its energy maximally spread everywhere on the sphere.

Diffusion Model

In the low b-value regime, the HARDI signal can be modeled as:

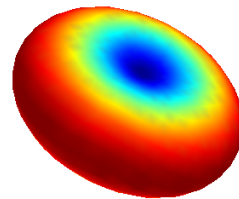$$S(\mathbf{u}) = \sum_i \alpha_i \exp(-b\mathbf{u}^T D_i \mathbf{u})$$
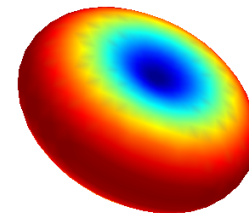
Spherical Ridgelets: Construction

- Use multi-resolution analysis on the sphere (similar to Freeden and Schreiner, 2008)

- The ridgelet generating function should be similar to a single fiber signal profile:

Single Fiber   **Signal Profile**   **Ridgelet Function**

Spherical Ridgelets: Construction

- Let $\kappa(x) = \exp(-\rho x(x+1))$ be a kernel function, which we use to define:

$$\kappa_j(x) = \kappa(2^{-j}x) = \exp\left\{-\rho\,\tfrac{x}{2^j}\left(\tfrac{x}{2^j}+1\right)\right\}, \quad j \in \mathbb{N}.$$

- The Gauss-Weierstrass kernel $\chi_{j,\mathbf{v}}(\mathbf{u})$ for resolution j and orientation v is defined as:

$$\chi_{j,\mathbf{v}}(\mathbf{u}) = \sum_{n=0}^{\infty} \tfrac{2n+1}{4\pi}\,\kappa_j(n)\,P_n(\mathbf{u}\cdot\mathbf{v}), \quad \forall \mathbf{u} \in \mathbb{S}^2,$$

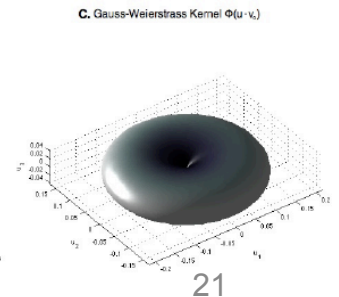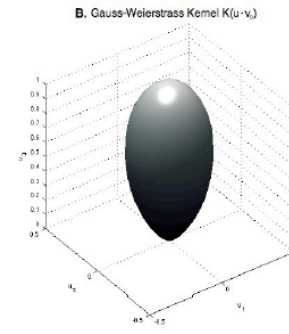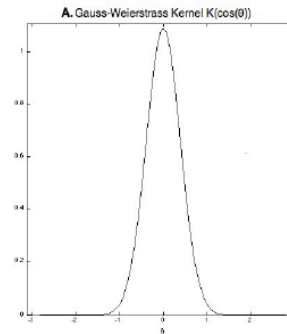where P_n is the legendre polynomial of order n.

Ridgelet Generating Function

Taking the Funk-Radon Transform of χ gives the Ridgelet Generating Function:

$$\hat{\chi}_{j,\mathbf{v}}(\mathbf{u}) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi}\, \kappa_j(n)\, \lambda_n\, P_n(\mathbf{u} \cdot \mathbf{v}), \quad \forall \mathbf{u} \in \mathbb{S}^2,$$

where,

$$\lambda_n = \begin{cases} 2\pi(-1)^{n/2} \frac{1 \cdot 3 \cdots (n-1)}{2 \cdot 4 \cdots n}, & \text{if } n \text{ is even} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$


A. Gauss-Weierstrass Kernel K(cos(θ))   B. Gauss-Weierstrass Kernel K(u·vₛ)   C. Gauss-Weierstrass Kernel Φ(u·vₛ)

Spherical Ridgelets

- Finally, the *spherical ridgelet function* for resolution j and direction v is computed using

$$\psi_{j,\mathbf{v}} = \hat{\chi}_{j+1,\mathbf{v}} - \hat{\chi}_{j,\mathbf{v}}, \; j \in \{-1, 0, 1, ...\}$$

- The semi-discrete set of spherical ridgelets $\{\psi_{j,\mathbf{v}}\}_{j \in \mathbb{N}, \mathbf{v} \in \mathbb{S}^2}$ is a frame for the subspace $\mathcal{S} \in \mathbb{L}(\mathbb{S}^2)$ of symmetric spherical functions.
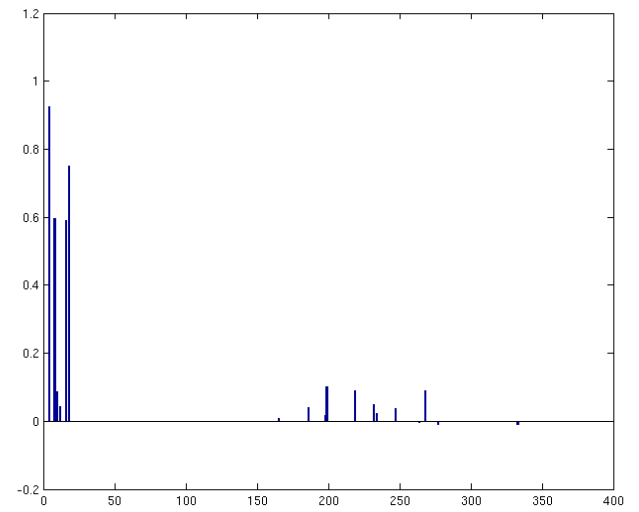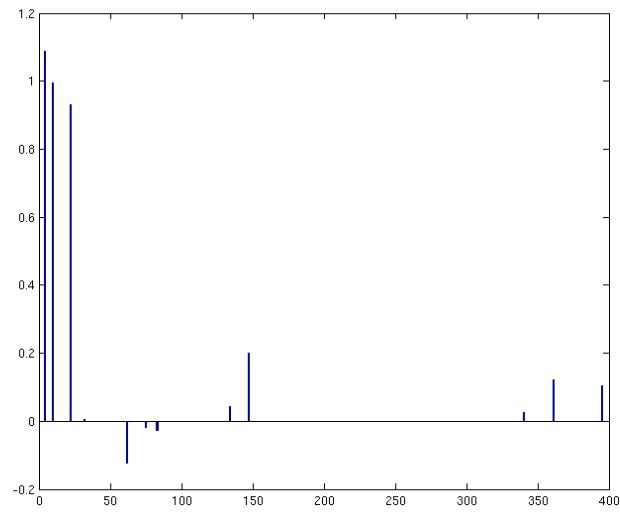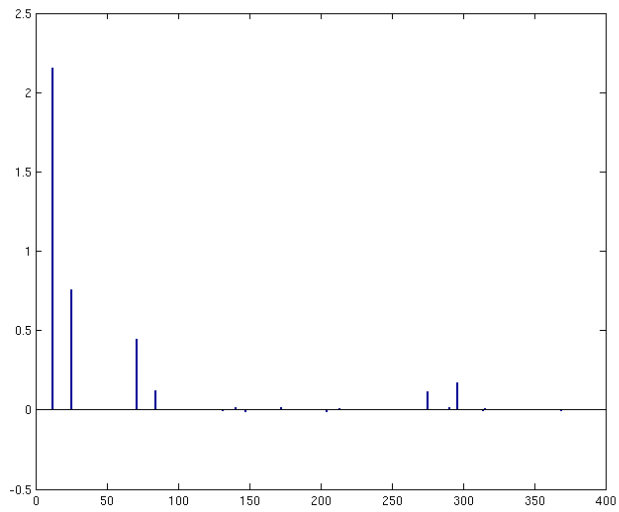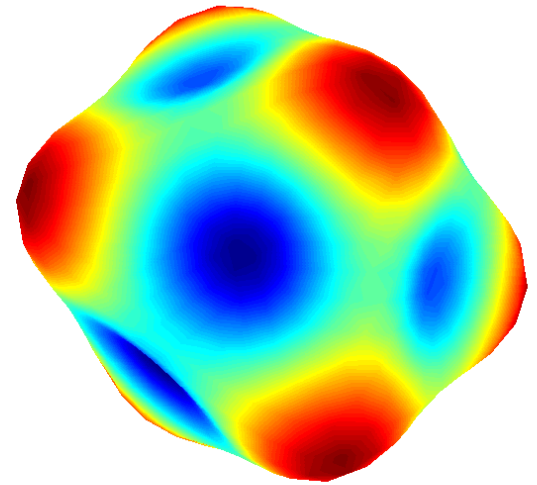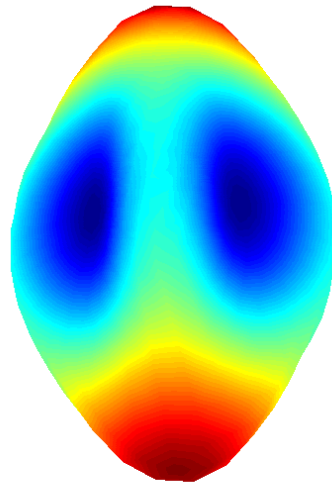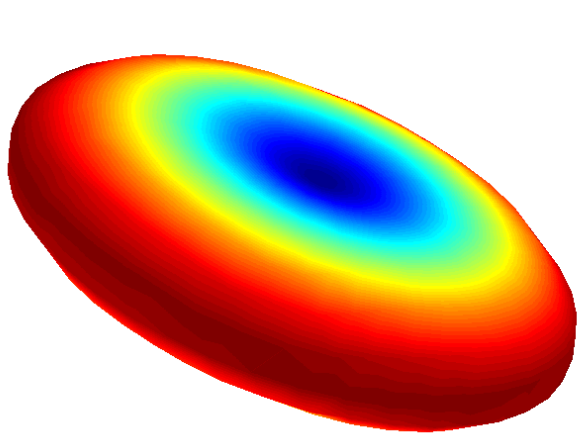
Spherical Ridgelets (SR) Vs Spherical Harmonics (SH)

SPHERICAL HARMONICS

SPHERICAL RIDGELETS
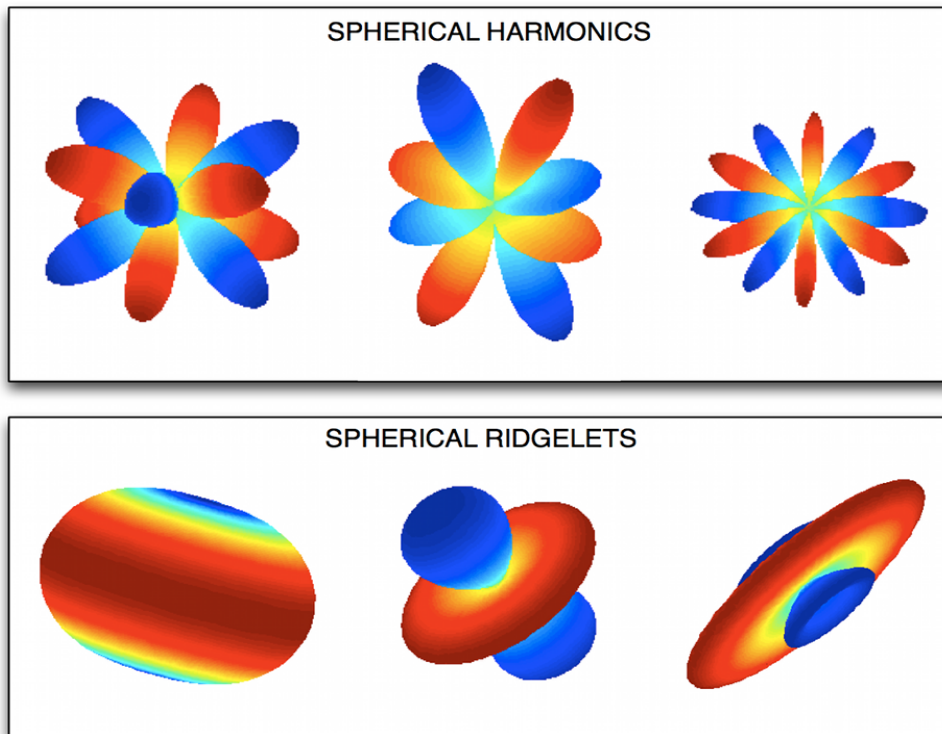
Central Property:

 As opposed to SH, the energy of the spherical ridgelets is concentrated alongside the great circles of $\mathbb{S}^2$

SR - properties

Implications

$$m \geq C.\mu^2(\Phi, \Psi).K. \log n$$
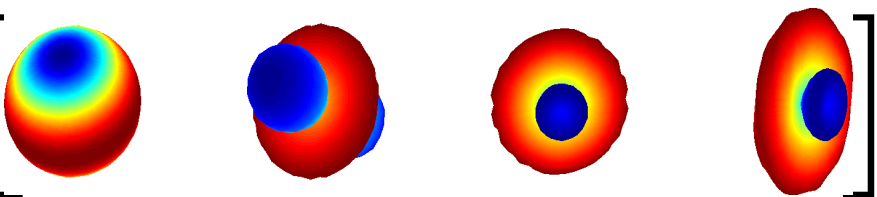
Note: m is quadratic in μ

μ for SH is about twice that of SR ➔ the minimum number of samples required by SH will be at-least 4 times that of SR !

Rule of thumb: we need at-most 5*K (sparsity factor) samples for accurate signal recovery (Romberg et al).

SR - Estimation

Given measurements in K diffusion
directions, the HARDI signal can be
represented as:

$$S = \begin{bmatrix} & & & \end{bmatrix} c + e$$

where, c is a vector of coefficients in SR
basis and e is noise. Our goal is to obtain
a sparse estimate of c.

SR - estimation

- Sparse estimation of c obtained using L1 minimization:

$$\min \|c\|_1 \; s.t., \|Ac - S\|_2^2 \leq \eta$$

- Several algorithms exist in the literature to solve this problem: Basis pursuit denoising, Fiesta, Nesta, L1-homotopy, weighed I1, etc.

  We use the method of Asif & Romberg et al 2010.

Composite Compressed Sensing

- Combine the sparseness constraint in the diffusion domain with a spatial regularity constraint in the spatial domain.

- We use the total-variation (TV) semi-norm, which has been widely used by the image processing community.

- For the kth gradient direction, the TV norm at location r is:

$$\|S_k(\mathbf{r})\|_{TV} = \sum_{x,y,z} \|\nabla S_k(\mathbf{r})\|$$

Composite CS

- The spatially constrained CS problem can now be defined as:

$$\min_c \left\{ \|c\|_1 + \mu\|Ac\|_{TV} \right\}$$

$$\text{s.t.} \quad \|Ac - S\|_2 \le \epsilon$$

or, using the Lagrangian formulation

$$\min_c \left\{ \frac{1}{2}\|Ac - s\|_2^2 + \lambda\|c\|_1 + \mu\|Ac\|_{TV}) \right\}.$$

Composite CS

This problem can be solved iteratively using the split Bregman iteration as follows:

$$\left(u^{t+1}, c^{t+1}\right) =$$

$$\arg\min_{c,u} \left\{ \tfrac{1}{2}\|u - S\|_2^2 + \lambda\|c\|_1 + \mu\|u\|_{TV} + \tfrac{\gamma}{2}\|u - Ac - p^t\|_2^2 \right\}$$

$$p^{t+1} = p^t + \left(Ac^{t+1} - u^{t+1}\right)$$

Composite CS

Splitting the variables, the solution can be easily obtained in 3 simple steps:
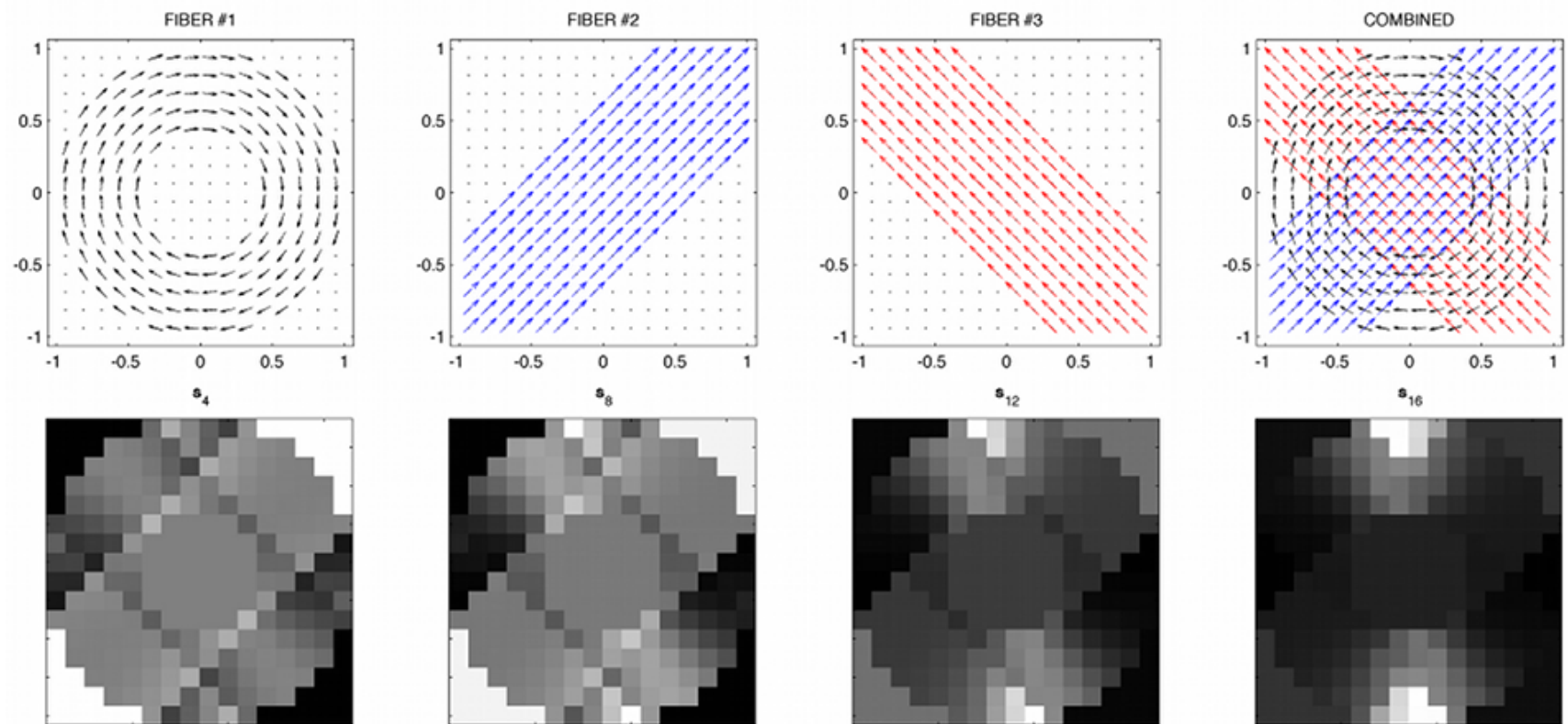
$$d^t = u^t - p^t$$

$$\text{Step 1: } c^{t+1} = \arg\min_c \left\{ \tfrac{1}{2} \|Ac - d^t\|_2^2 + \alpha \|c\|_1 \right\}$$

$$\text{Step 2: } u^{t+1} = \arg\min_u \left\{ \tfrac{1}{2} \|u - d^t\|_2^2 + \beta \|u\|_{TV} \right\}$$

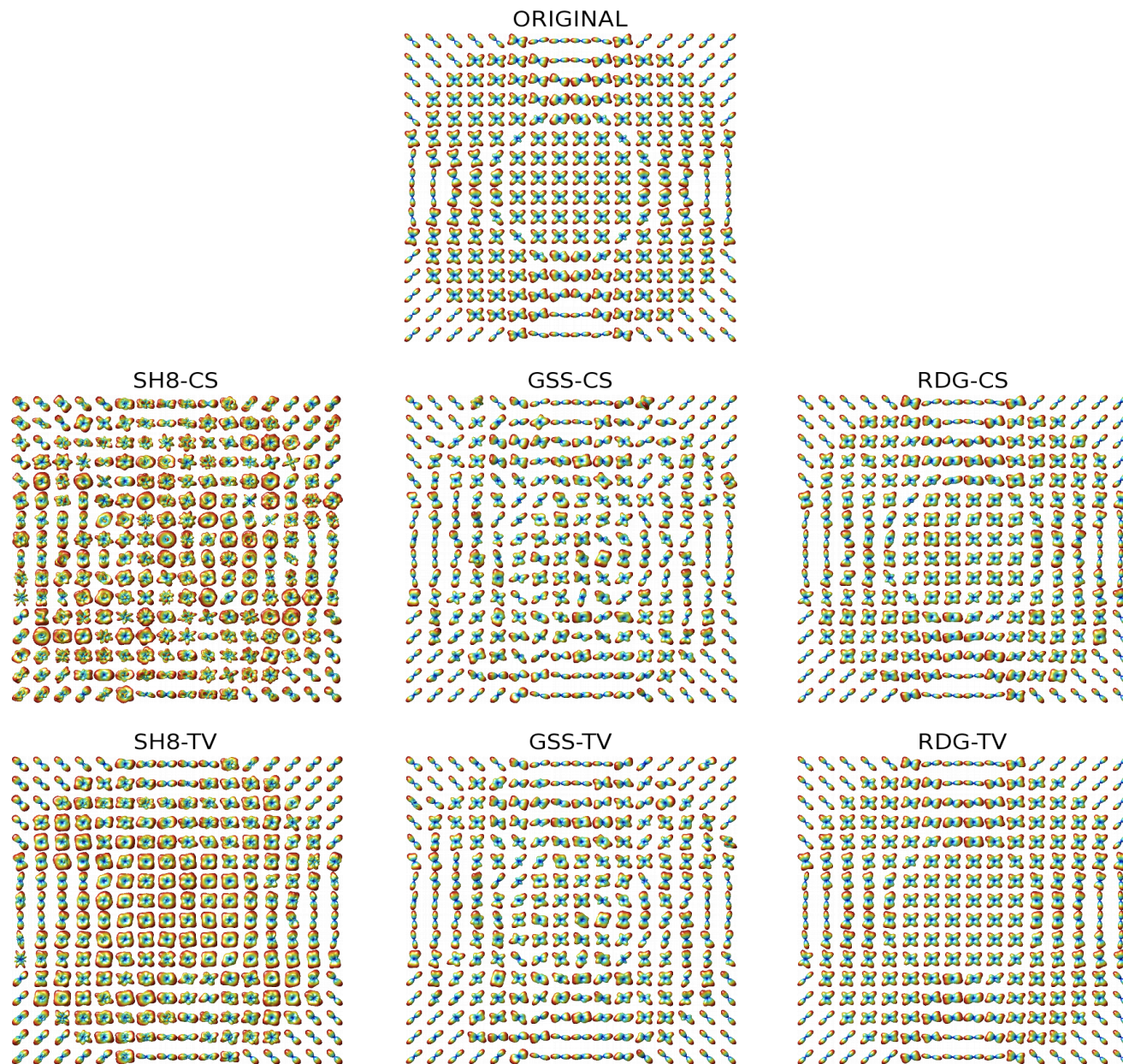$$\text{Step 3: } p^{t+1} = p^t + \left(Ac^{t+1} - u^{t+1} \right)$$

where, we start with
$$u^0 = S$$
$$p^0 = 0$$

Experiments

- Up to three ``fibers" per voxel

$$K \in [16, 32],\ b = \{1000, 3000\} s/mm^2, \text{SNR} \in [5, 40]$$

ORIGINAL

SH8-CS   GSS-CS   RDG-CS

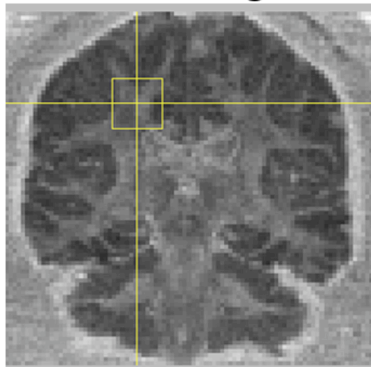SH8-TV   GSS-TV   RDG-TV

SH8 – spherical harmonics, GSS – Gaussian mixtures, RDG – Ridgelets,
TV – total variation, CS – pure compressed sensing , K=16, SNR = 18dB
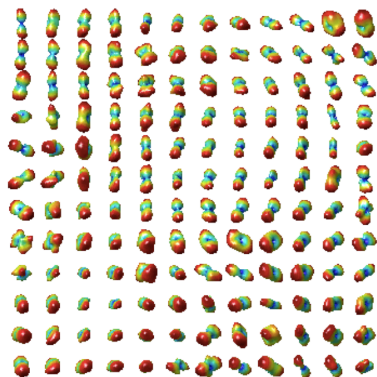
Experiments
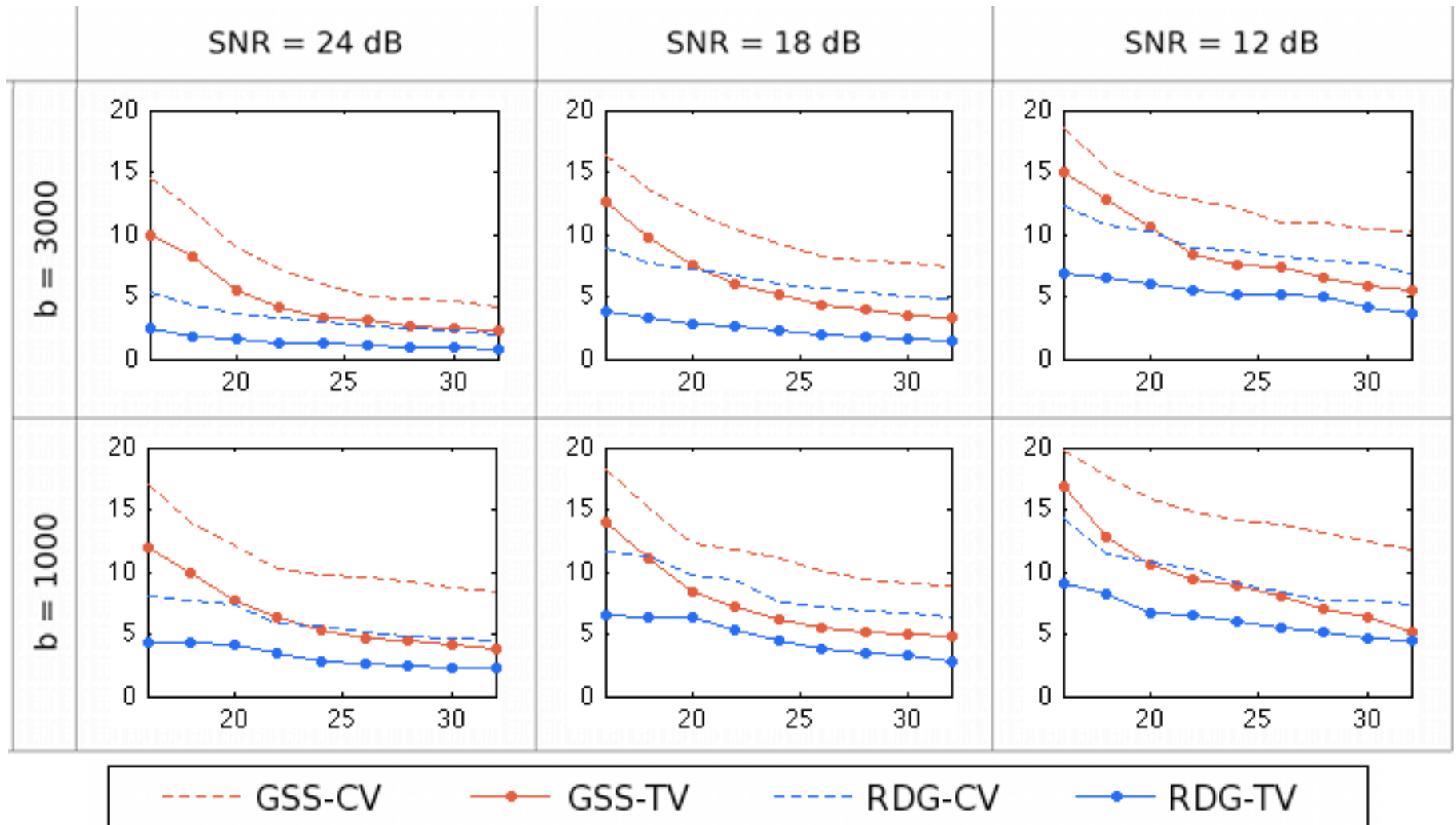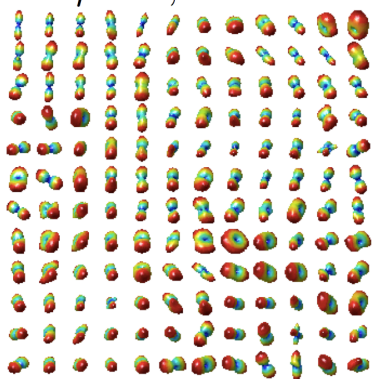

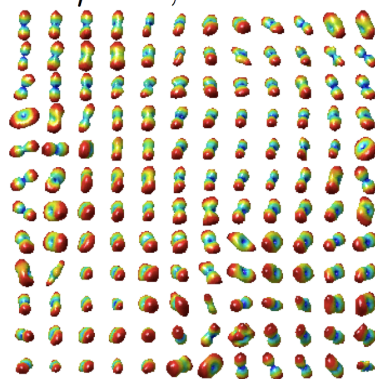Angular error in degrees as a function of K
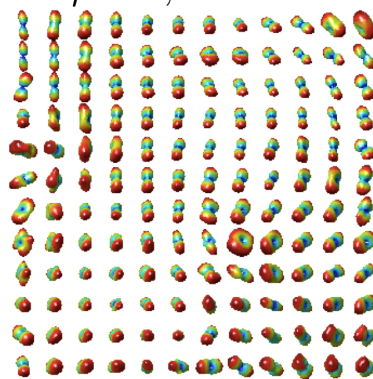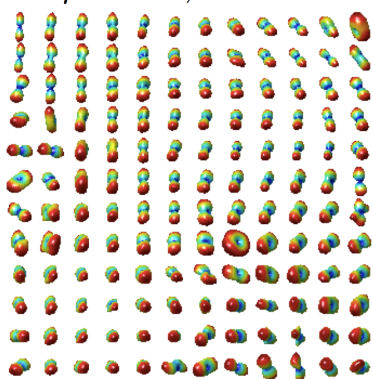
34

GFA image
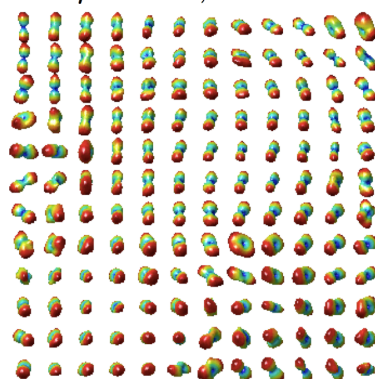
$\mu = 0, K = 51$
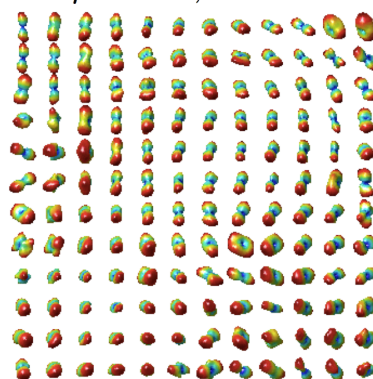
$\mu = 0, K = 16$

$\mu = 0, K = 24$

$\mu = 0, K = 32$

$\mu = 0.3, K = 16$

$\mu = 0.3, K = 24$

$\mu = 0.3, K = 32$

|	$K = 16$	$K = 24$	$K = 32$
RDG-CS	9.11 ± 2.23	5.31 ± 1.23	3.74 ± 0.87
RDG-TV	1.78 ± 0.37	0.92 ± 0.21	0.24 ± 0.06

Extension to multi-shell data

- Diffusion Spectrum Imaging (DSI) allows to capture the average ensemble diffusion propagator – EAP (displacement probability).

    - Requires a large number of measurements and hence very time consuming

- Several authors have proposed alternatives to DSI: Descoteaux et al, 2010 – DPI, Wu & Alexander et al 2007– HYDI.

    - Main idea is to recover the EAP from data sampled on a few concentric shells i.e., multiple b-value

- Recent work has focused on compressed sensing based signal recovery in the entire q-space (Merlet et al, Cheng et al, Gramfort et al, Rathi et al).

Challenges

- Noise contamination becomes more acute
- Signal decay is no longer exponential, rather it becomes bi-exponential.
- Allows to compute the mean-squared displacement and return-to-origin probability measures – possibly more sensitive to cellular changes (Assaf 2005, Cheung 2009)
- Signal decay with q-value might be correlated with axonal size (?)

Single to Multiple Shells
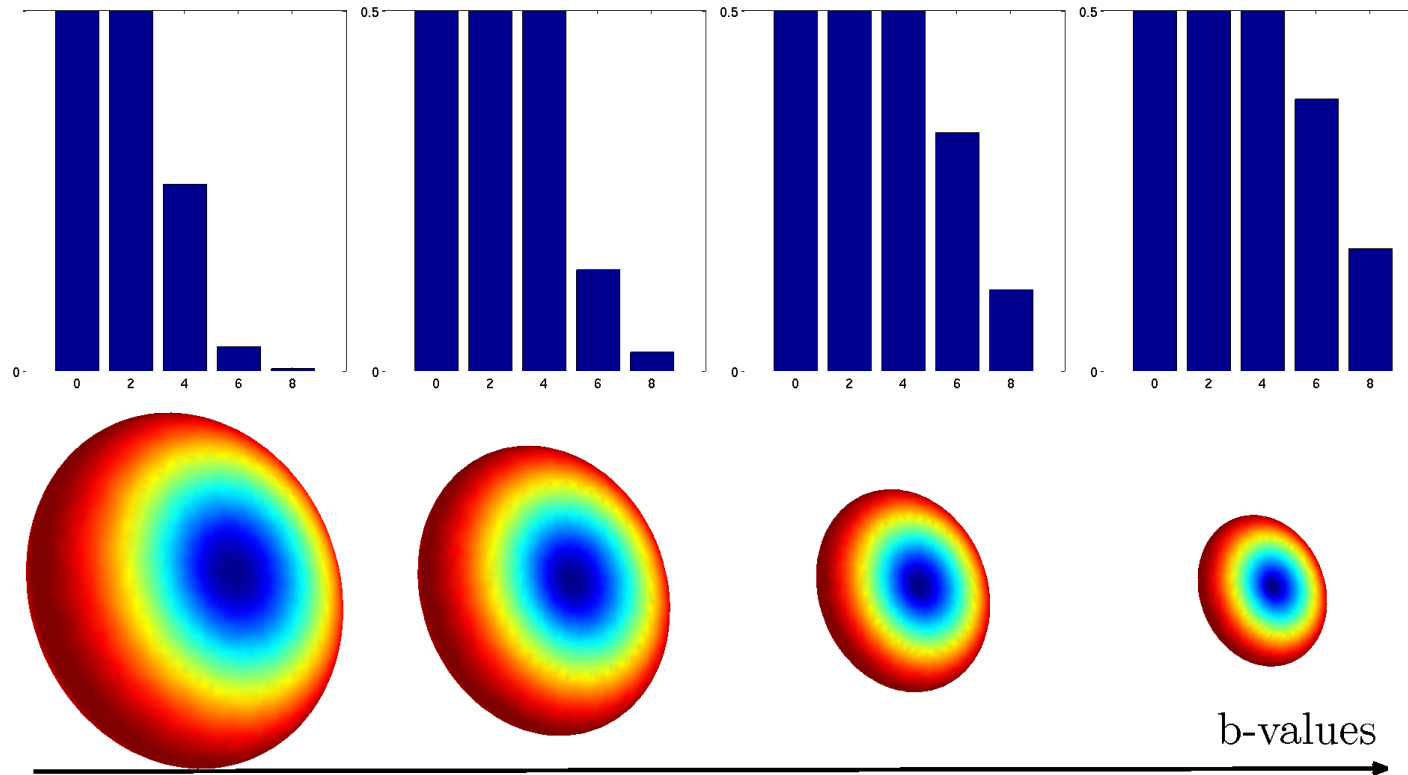
Top row: Strength of the order of spherical harmonics required to represent signal with increasing b-value. Also notice the decrease in magnitude.

Observations

- Radial decay (with b-value) of the signal is bi-exponential (i.e. monotonically decreasing) [Mulkern et al].

- High frequency components needed to represent data at higher b-value.

Spherical Ridgelets for multi-shell

Key concept:

1. Use SR to model spherical data for all shells (consistency in spherical domain)
2. Use a radial decay term to model decay of signal with b-value.

Challenges

- Independent estimation of spherical domain data at each b-value shell will provide inconsistent results due to increasing noise.

- Independent estimation of the radial term for each gradient direction will provide inconsistent results in the spherical domain.

Consensus building

- Ensure consistency between the multiple b-value shells (radial term) and spherical domain by designing an appropriate cost function.
- Use ADMM to solve this complex optimization problem.

$$\min_{c_i}\{\| c_i \|_1 + \mu \| Ac_i \|_{TV}\}$$

$$\text{s. t. } \| Ac_i - S_i \|_2 \leq \epsilon_1, \quad i = \{1, 2, ..n_b\}$$

$$\text{s. t. } \| \Theta(a_j, k_j) - S^j \|_2 \leq \epsilon_2 \quad j = \{1, 2, \ldots N\}$$

n_b = number of b-value shells
N = number of gradient directions per shell.

Radial Term

- We propose a variant of the CDF of the Burr distribution to model radial decay.

$$\Theta(a, k) = (1 + x^a)^{-k}$$

- Monotonically decreasing function
- Ranges strictly between [0,1].
- Only 2 parameters to estimate.
- Can model bi-exponential decay.

Experiments
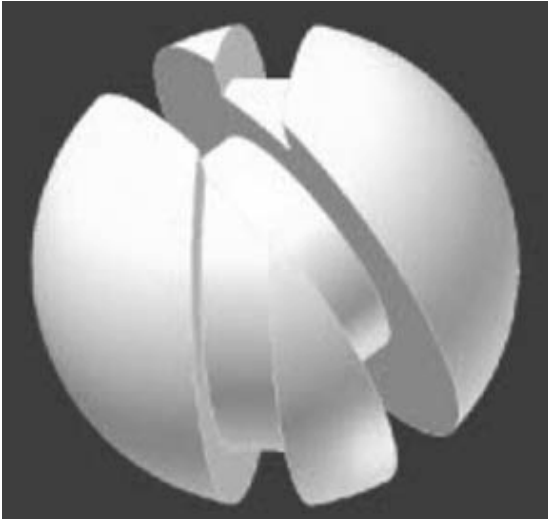
1. Built a physical phantom (F. Laun et. al).
2. Synthetic polyfil fibers (15μm).
3. Angular crossing of 45 degrees.
4. Gold standard data acquired with 81 gradient directions and 5 b-values (1000,2000,3000,4000,5000).
5. 10 repetitions acquired to obtain an average data set that forms the "gold standard".

6. Test data set: 5 sets each of the following were acquired :
    1. gradient directions = {16,20,24,26,30,36,42,60,81}.
    2. For each of the following b-values: {1000,2000,3000,4000,5000}.
    3. Average SNR over all directions is 8.5

Experiments

- Computed the following error to assess the quality of signal reconstruction

  - Angular error

  - Incorrect percent of peaks detected

  - NMSE in signal reconstruction (gold std)

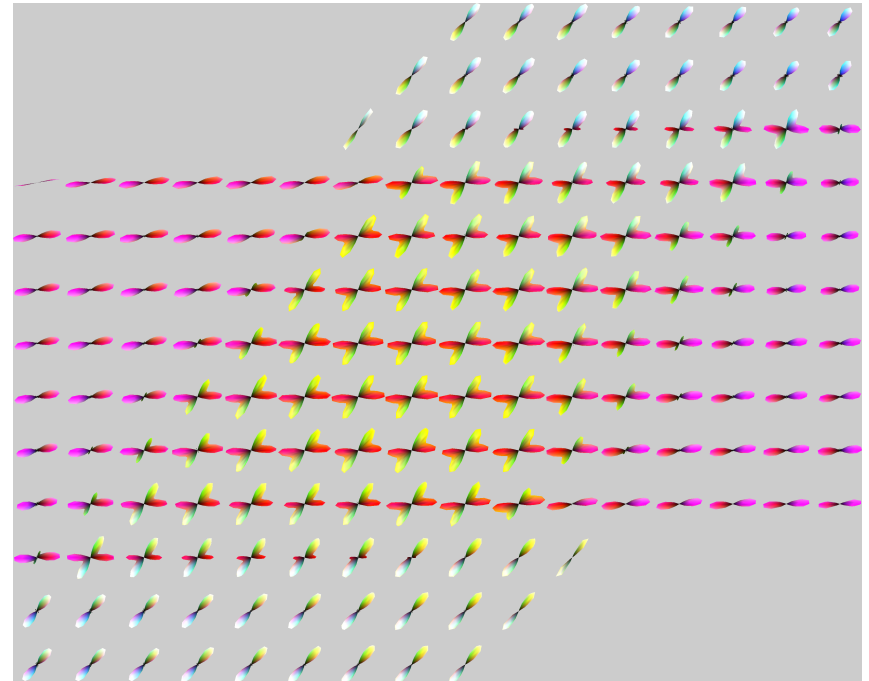 - NMSE in estimation of return-to-origin probability
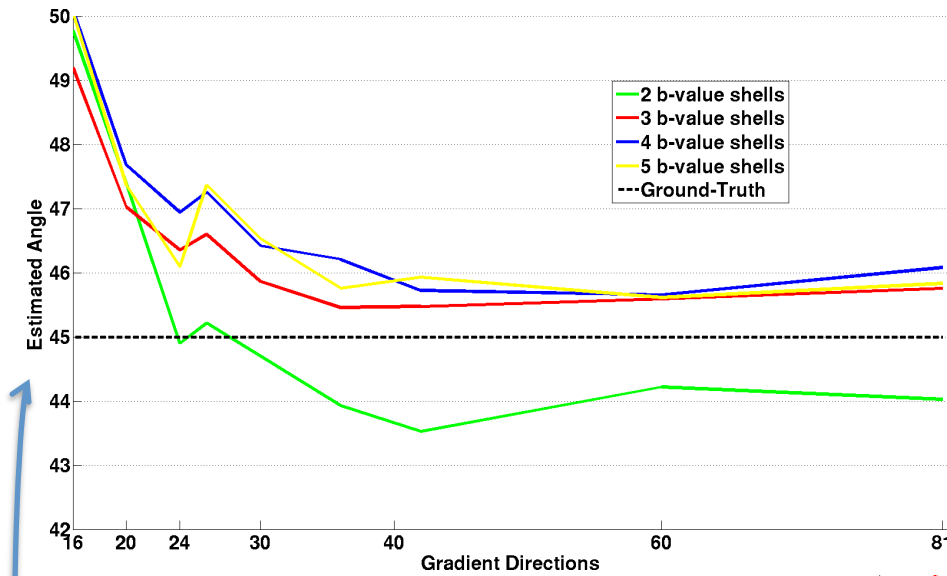
Results
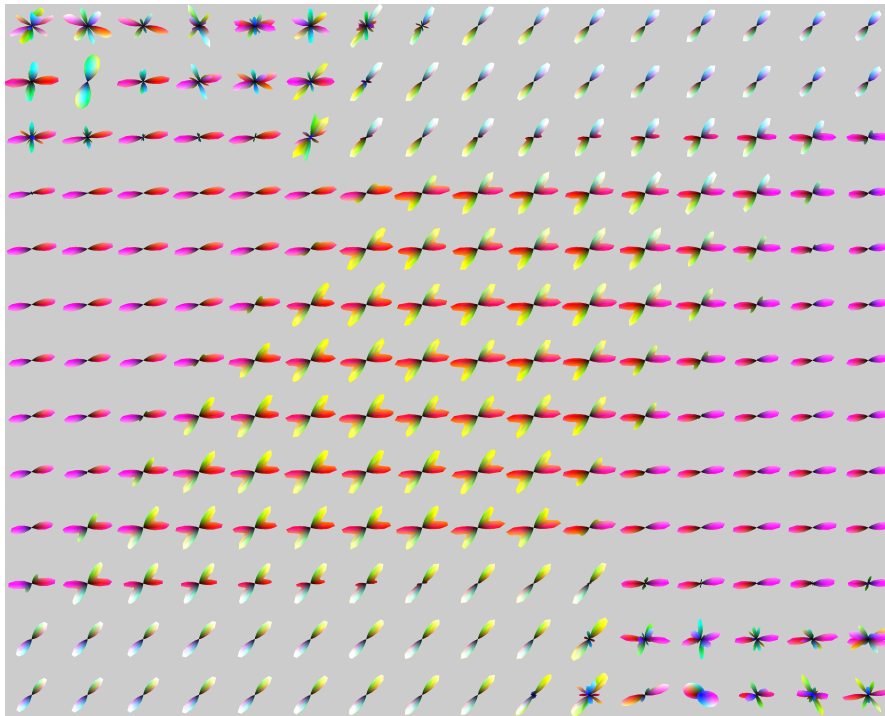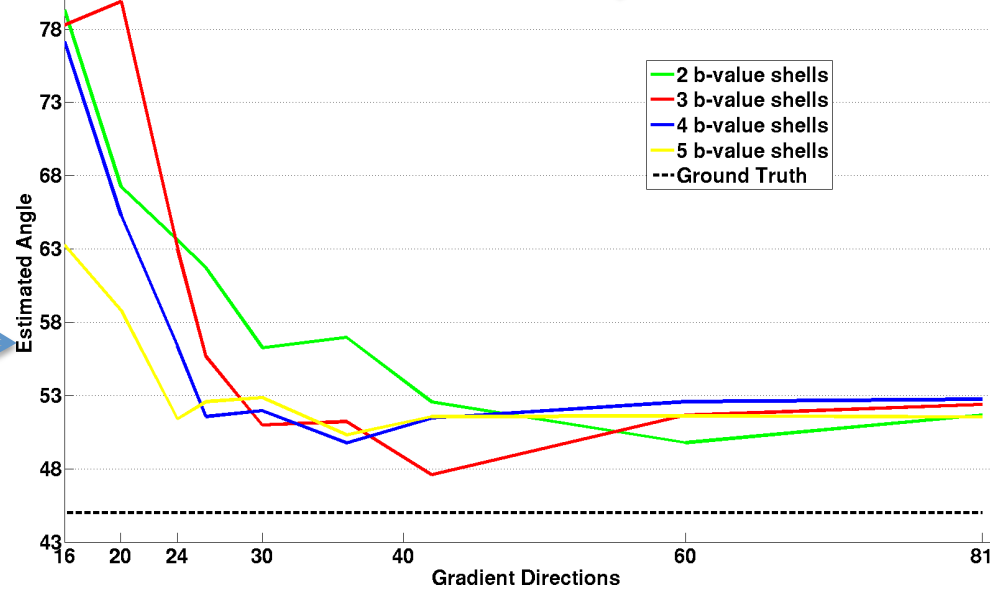

SHORE "Gold Standard"

SR   "Gold Standard"

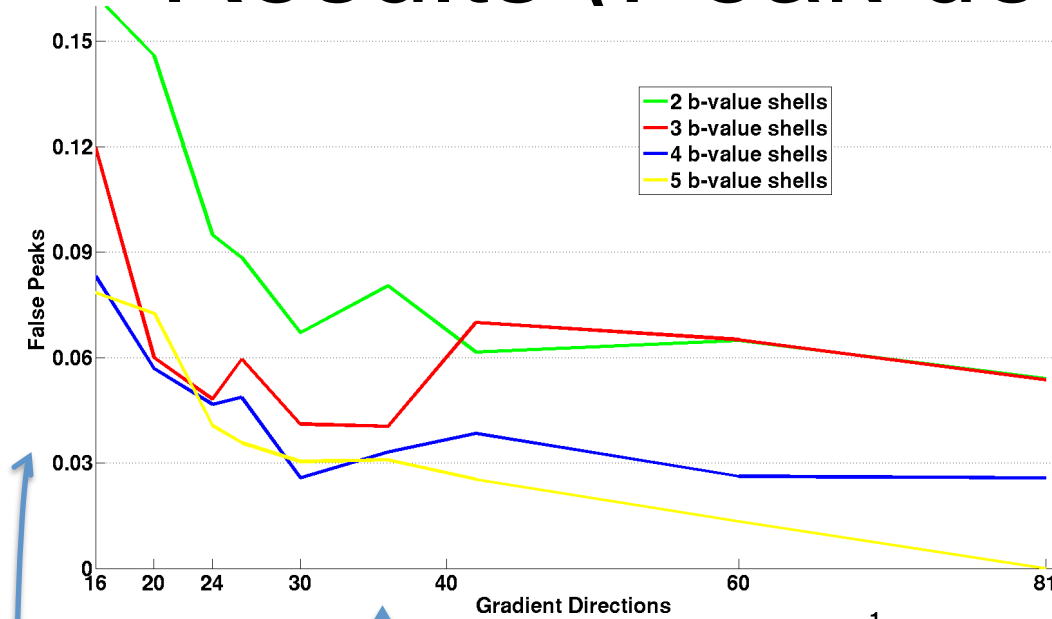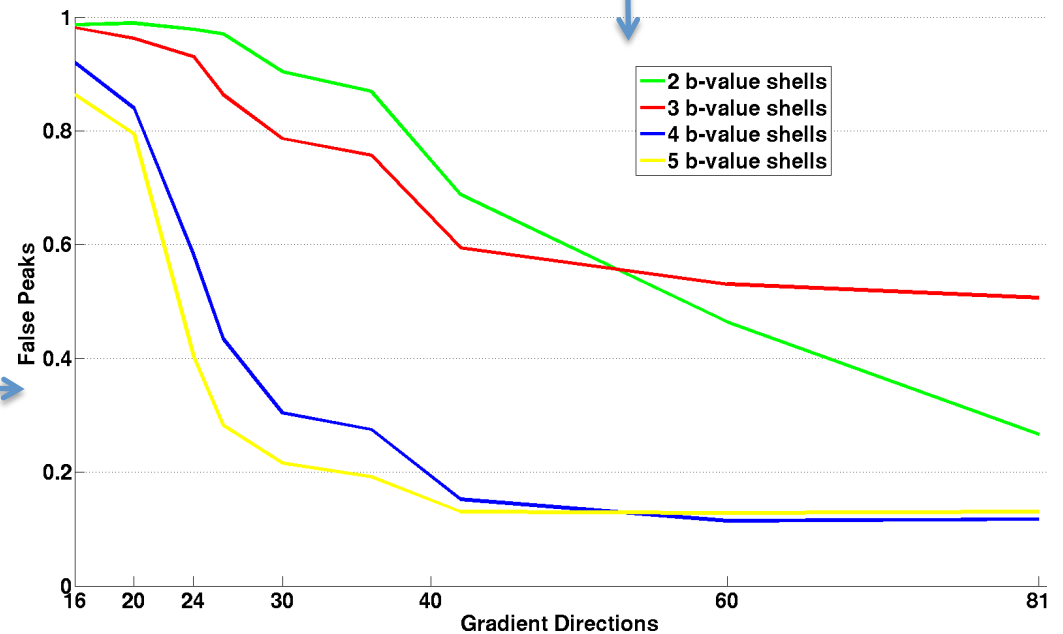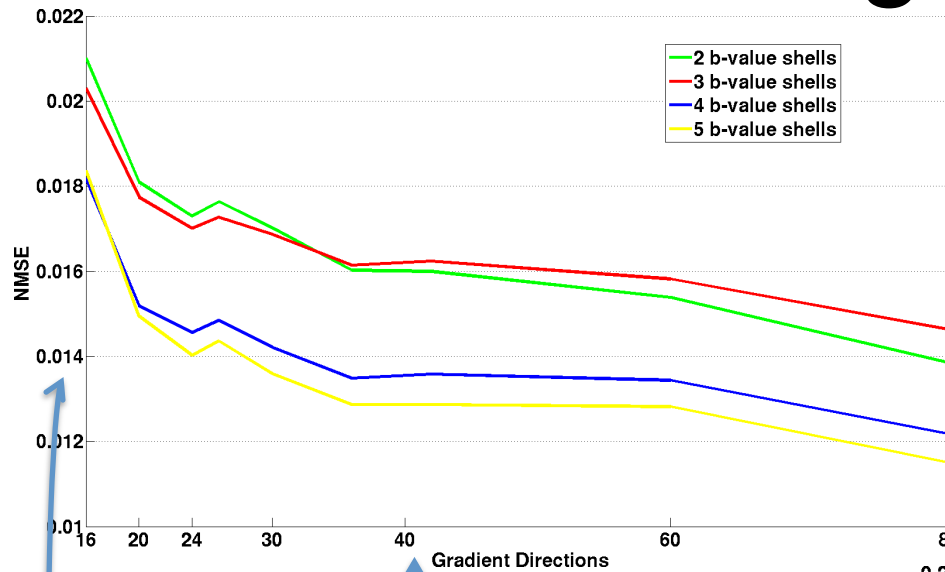# Results

Results (Peak detection error)
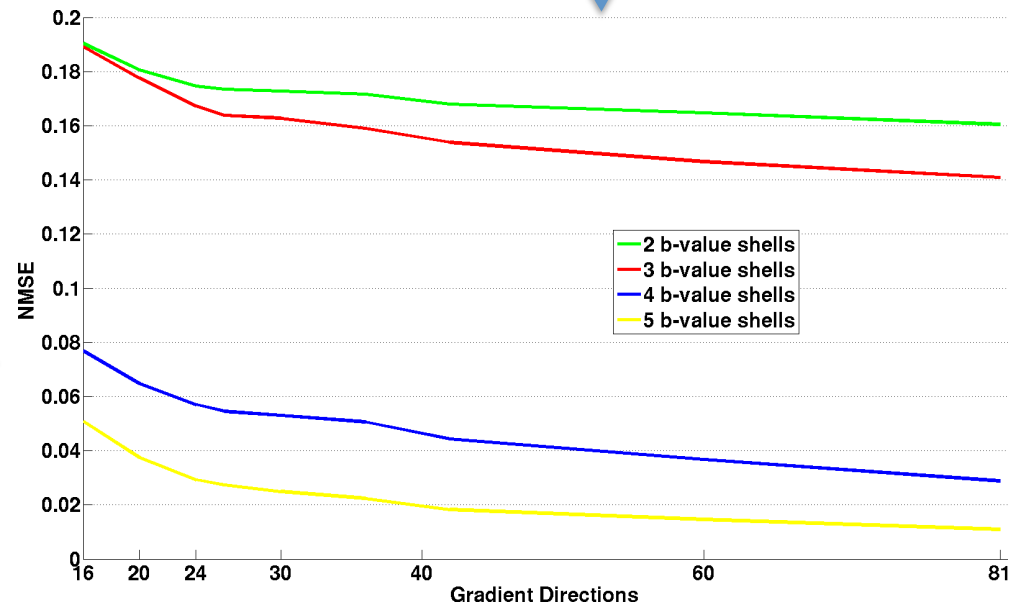

SR

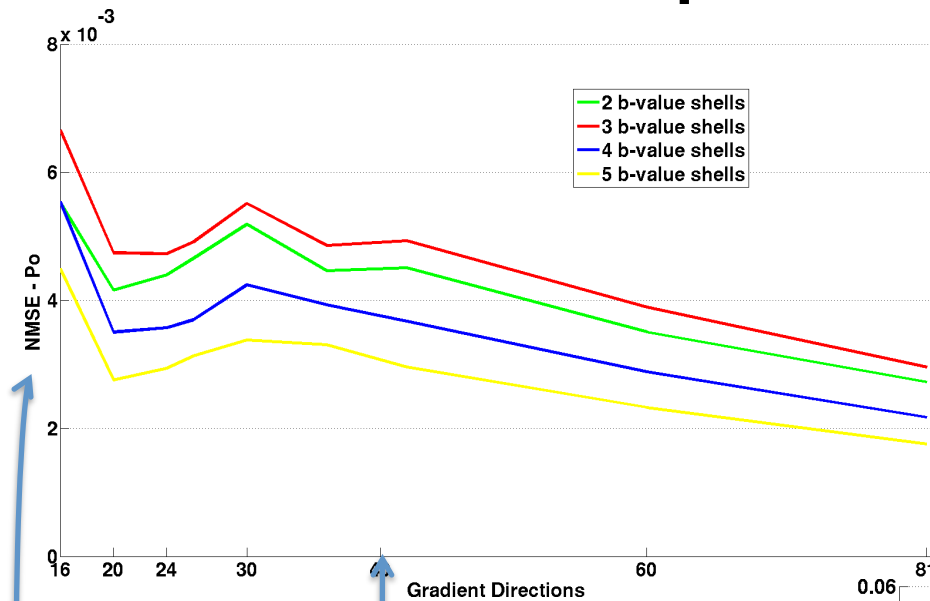SHORE

Notice the range
of error.

NMSE in signal estimation

SHORE

SR

Notice the range
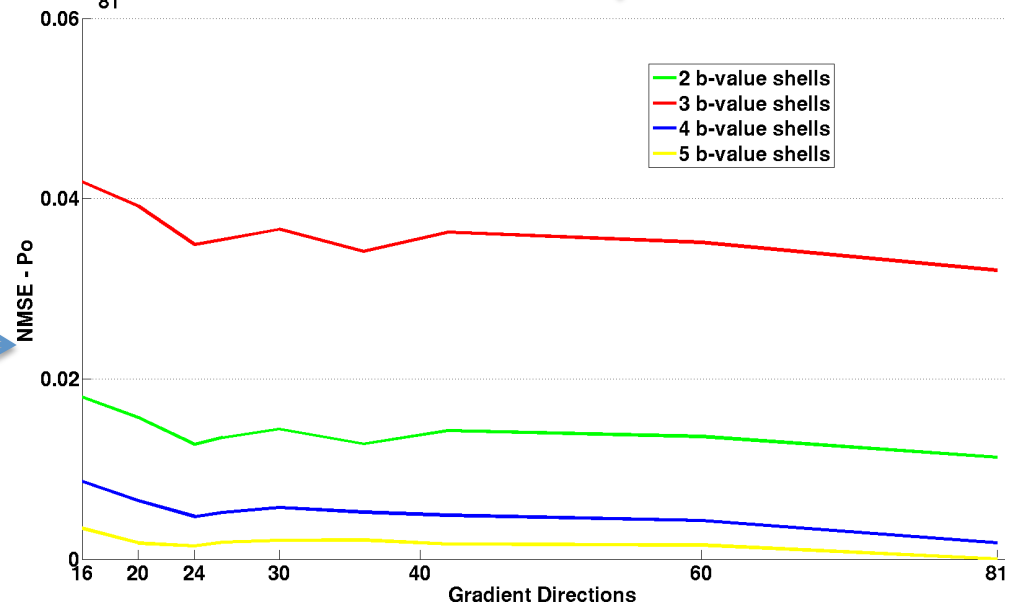of error.

NMSE in Po (return to origin probability)

SHORE
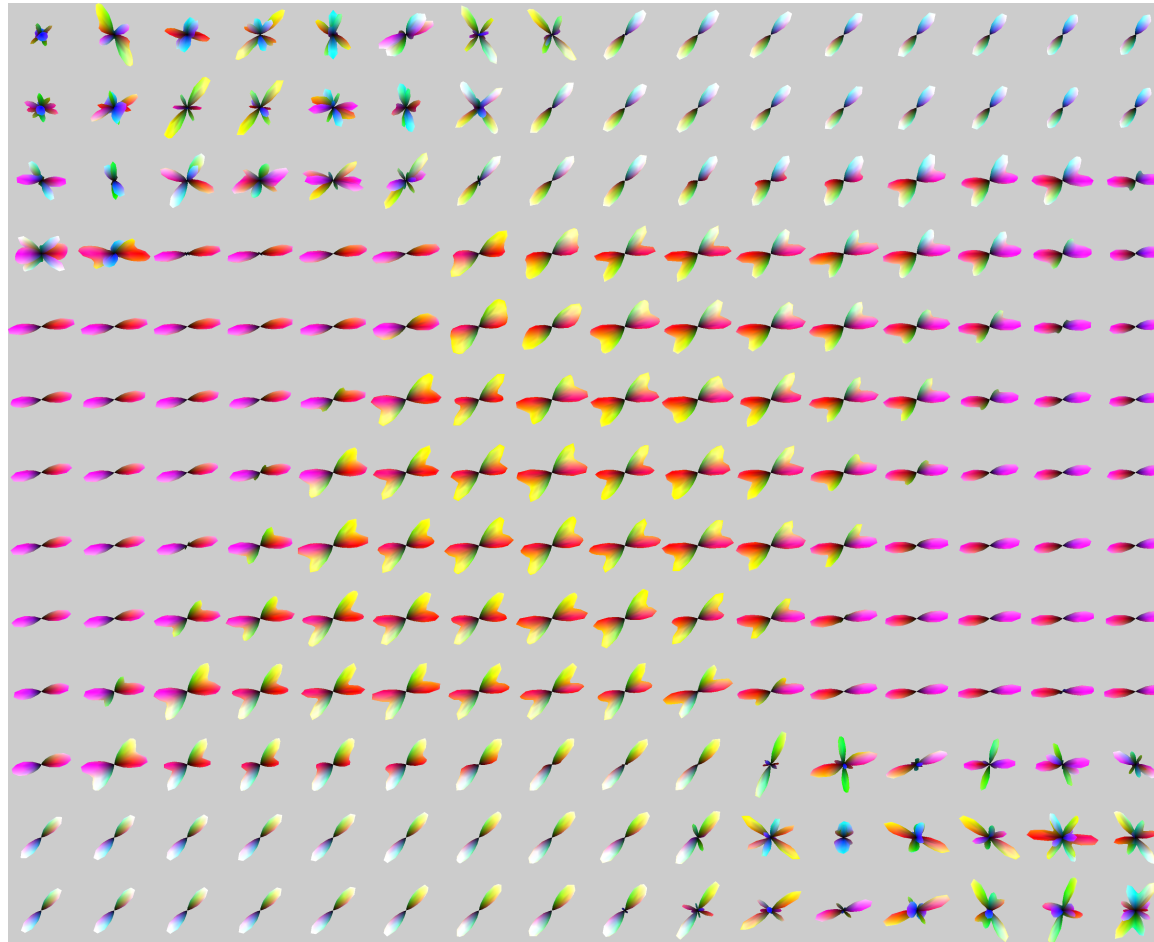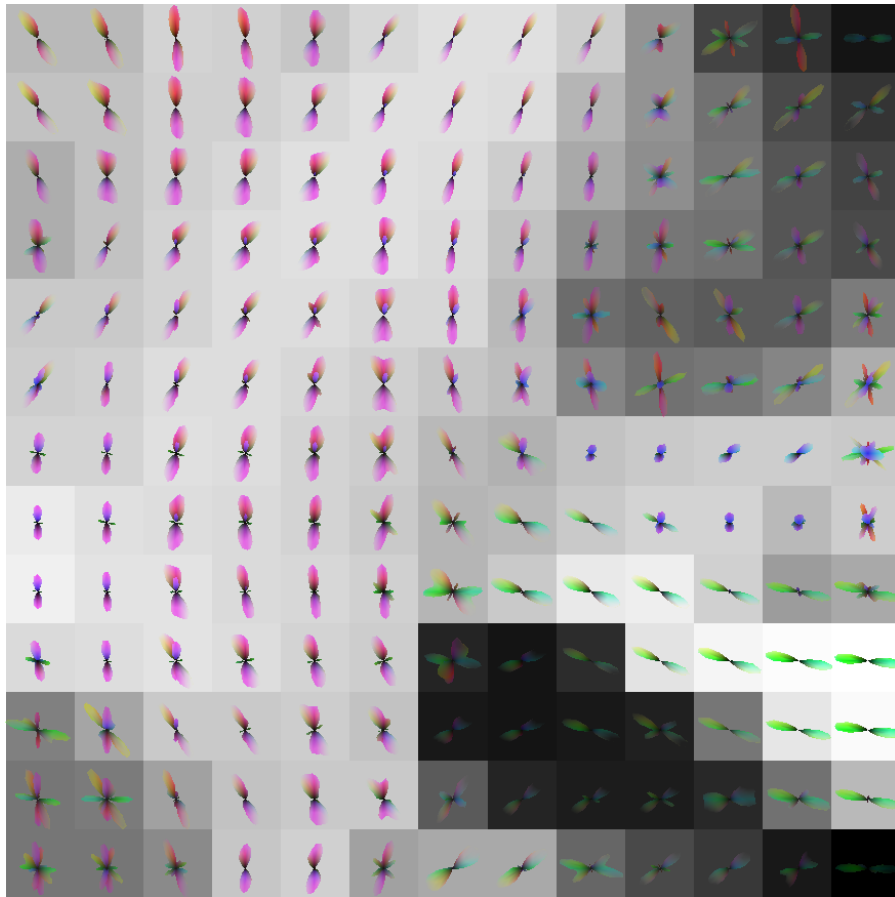
SR

Notice the range
of error.

Colorful figures

K = 20, and #b-shells = 3

Human in-vivo data

- Data acquired on a 3T scanner
- 2.5 x 2.5 x 2.5 mm^3 spatial resolution
- 60 gradient directions per shell
- 4 shells of {900,2000,3600,5600}.
  - For testing purposes, we downsampled the data to obtain the required number of gradient directions and used specific b-value shells.
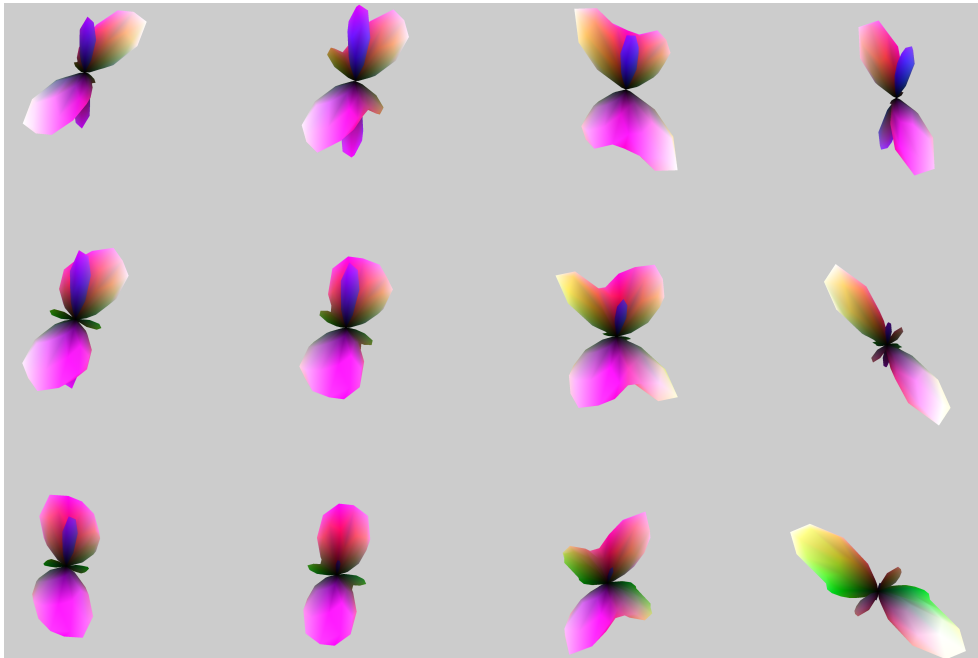
In-vivo data
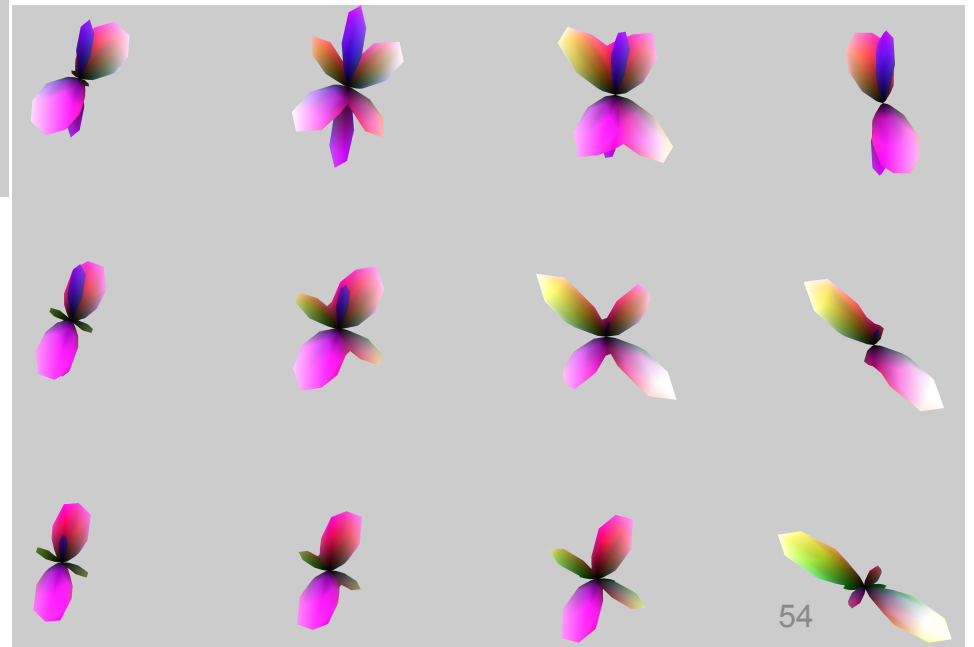

Gold Standard
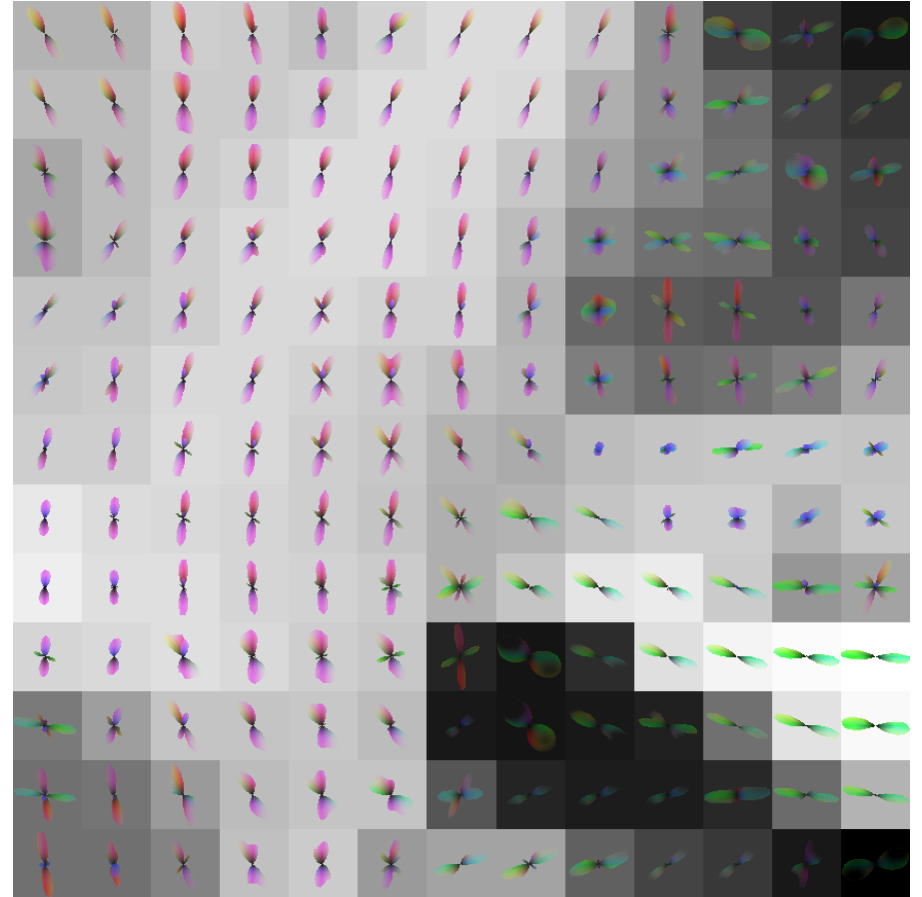
K=20, #b shells=3
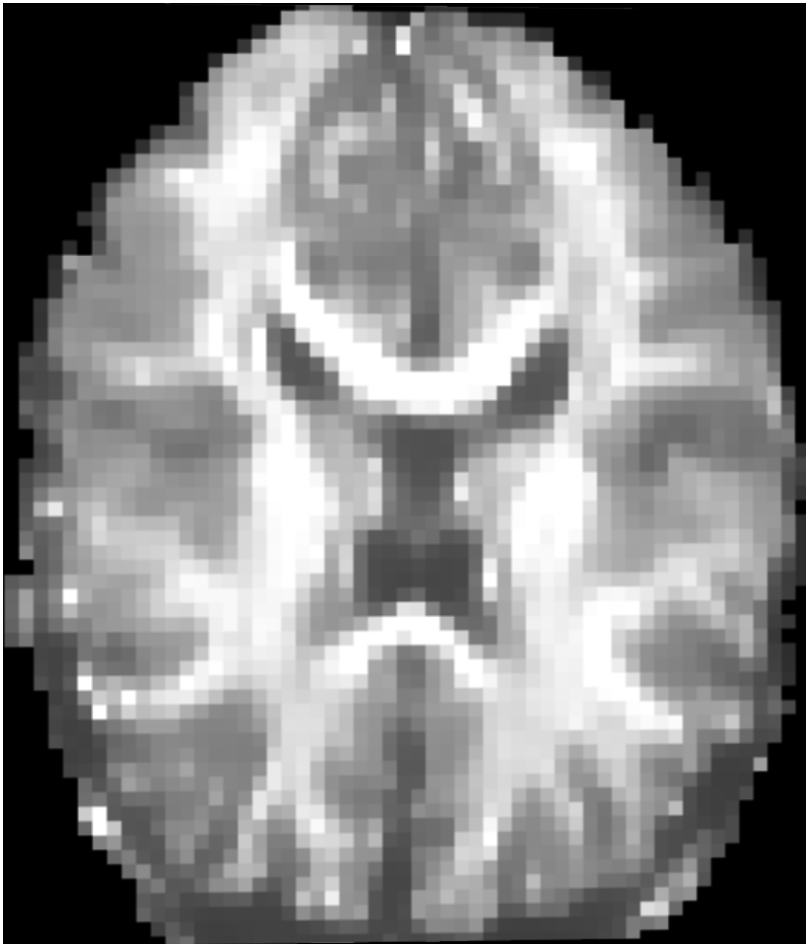
Angular error of 3.4 degrees

In-vivo data (zoomed)

Gold standard

K=20, #b-shells =3

In-vivo results

Po image
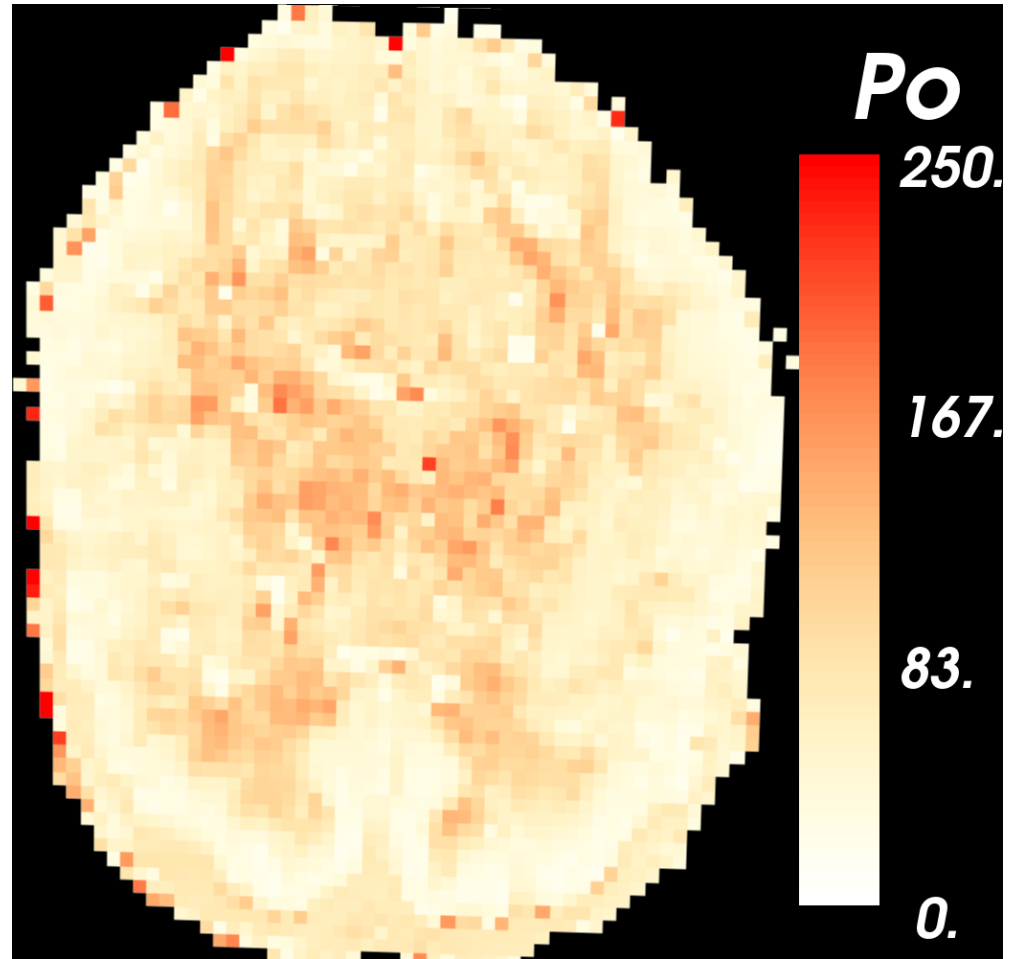
Error in estimation of Po (scaled)

Rathi 2014, MeDIA.

Conclusion

- We proposed a technique for signal recovery in the entire q-space.

- 3 b-value shells and [20,24] directions per shell give satisfactory results.

- Any type of data analysis model can be used for further processing (fast, slow diffusing fractions, compute EAP, Kurtosis model, fiber-ODF, etc.).

Limitations

- No analytical expression for direct computation of the diffusion propagator.

- Effect of error on long-range tracts is not known.

Thank you for your attention

Model based Sparse reconstruction

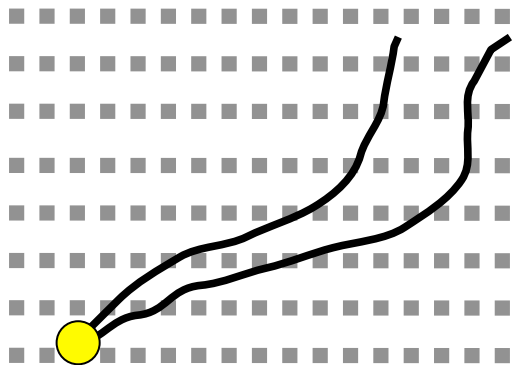# Combine Tractography and model estimation

- Tractography – used to trace white matter fiber bundles connecting cortical and sub-cortical gray matter areas.

- Used in most neuroscience studies to understand white matter connectivity and related pathologies.

- Why not combine the model estimation and tractography ?

- Provide estimates for diffusion measures such as FA while tracing fiber tracts.

We want to trace a fiber
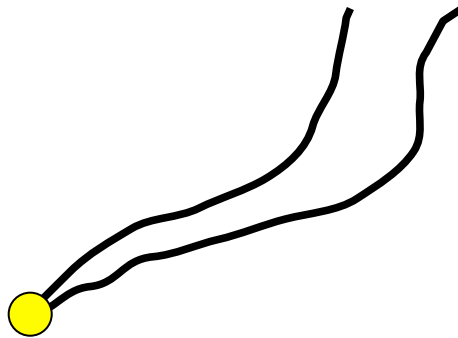
independent process

causal process

unscented Kalman filter (UKF)

neural fibers $\xrightarrow{x_t}$ measured signal $\xrightarrow{y_t}$

Unscented Kalman Filter

estimated model $\xrightarrow{\hat{x}_t}$ sigma points $\xrightarrow{\bar{x}_{t+1|t}}$ predicted signal

$\hat{x}_{t+1}$

$\bar{y}_{t+1|t}$

Malcolm et al. 2010
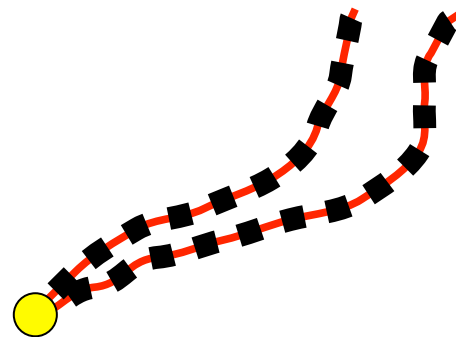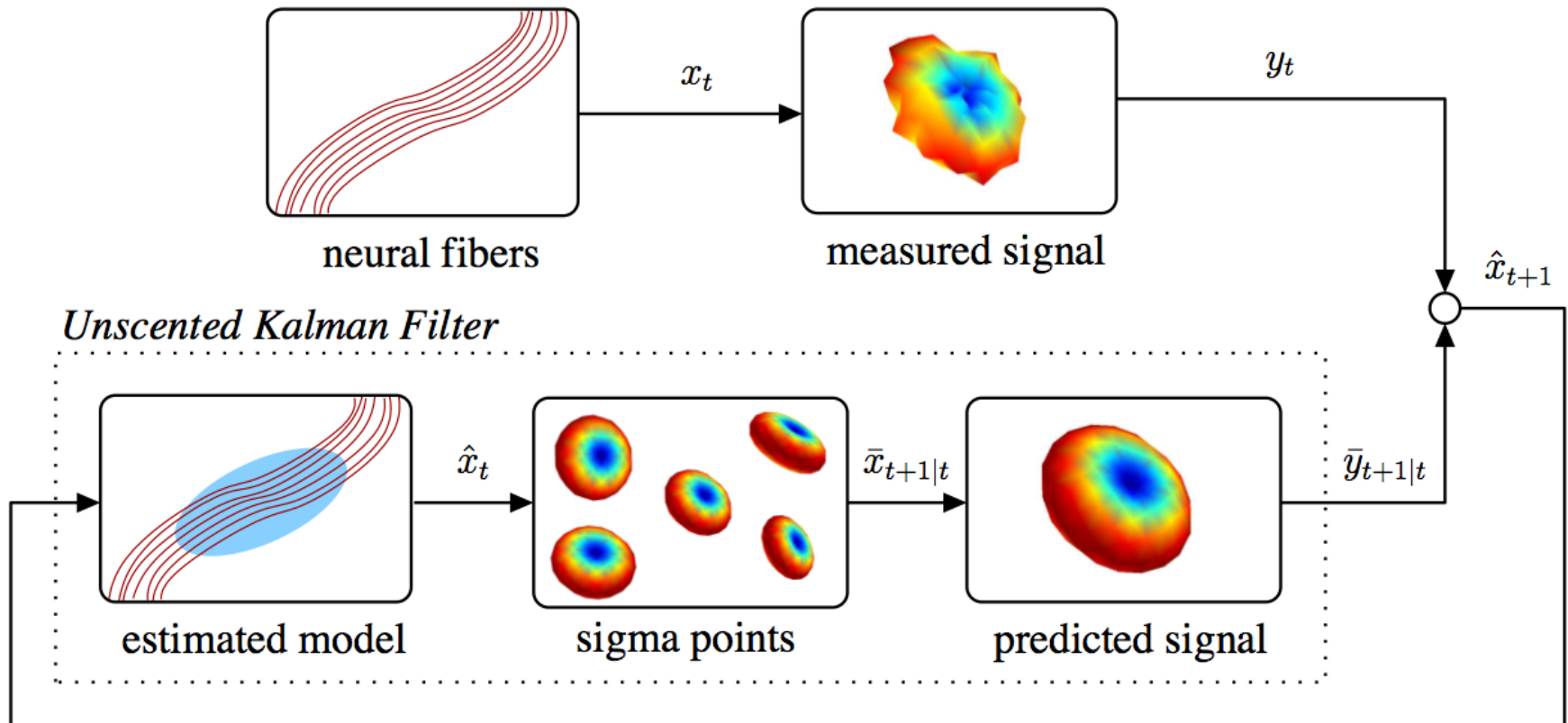
UKF for estimating diffusion propagator

Propose a novel bi-exponential multi-tensor model for representing diffusion signal.

a). Bi-exponential – captures the radial decay in the signal for high b-values.

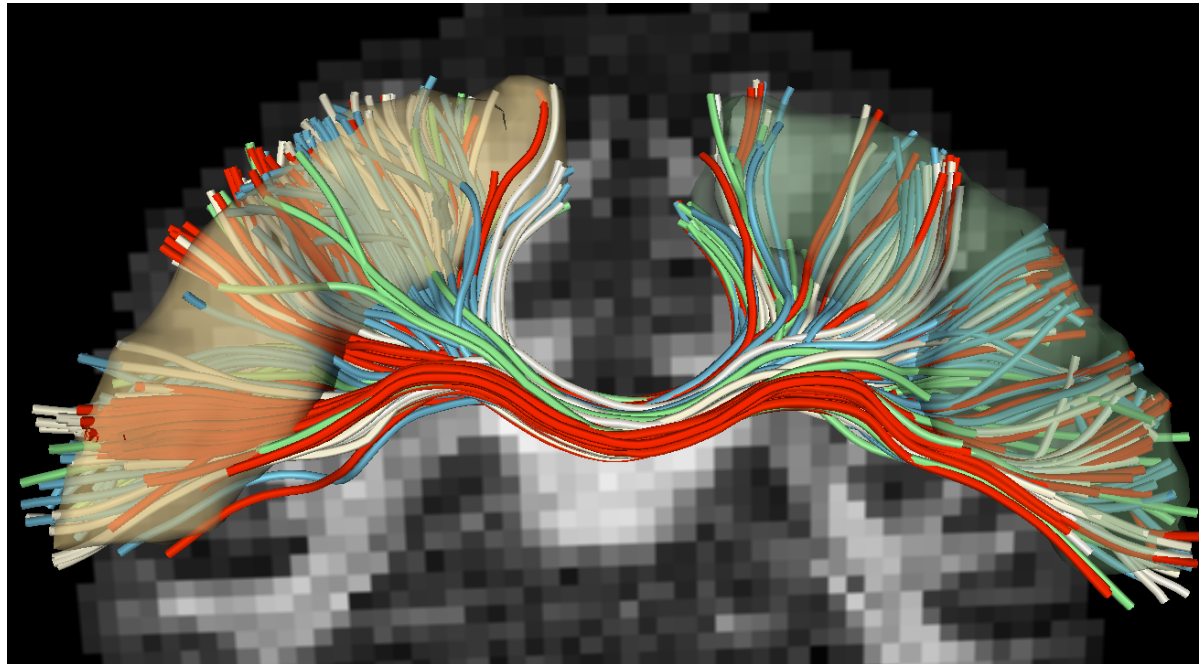b). multi-tensor – captures multi-fiber crossing or kissing configurations.

$$S(b, \mathbf{u}) = \frac{1}{n} \sum_{i=1}^{n} w \exp(-b\mathbf{u}^T D_i \mathbf{u}) + (1 - w) \exp(-b\mathbf{u}^T \bar{D}_i \mathbf{u}).$$

# UKF for estimating diffusion propagator
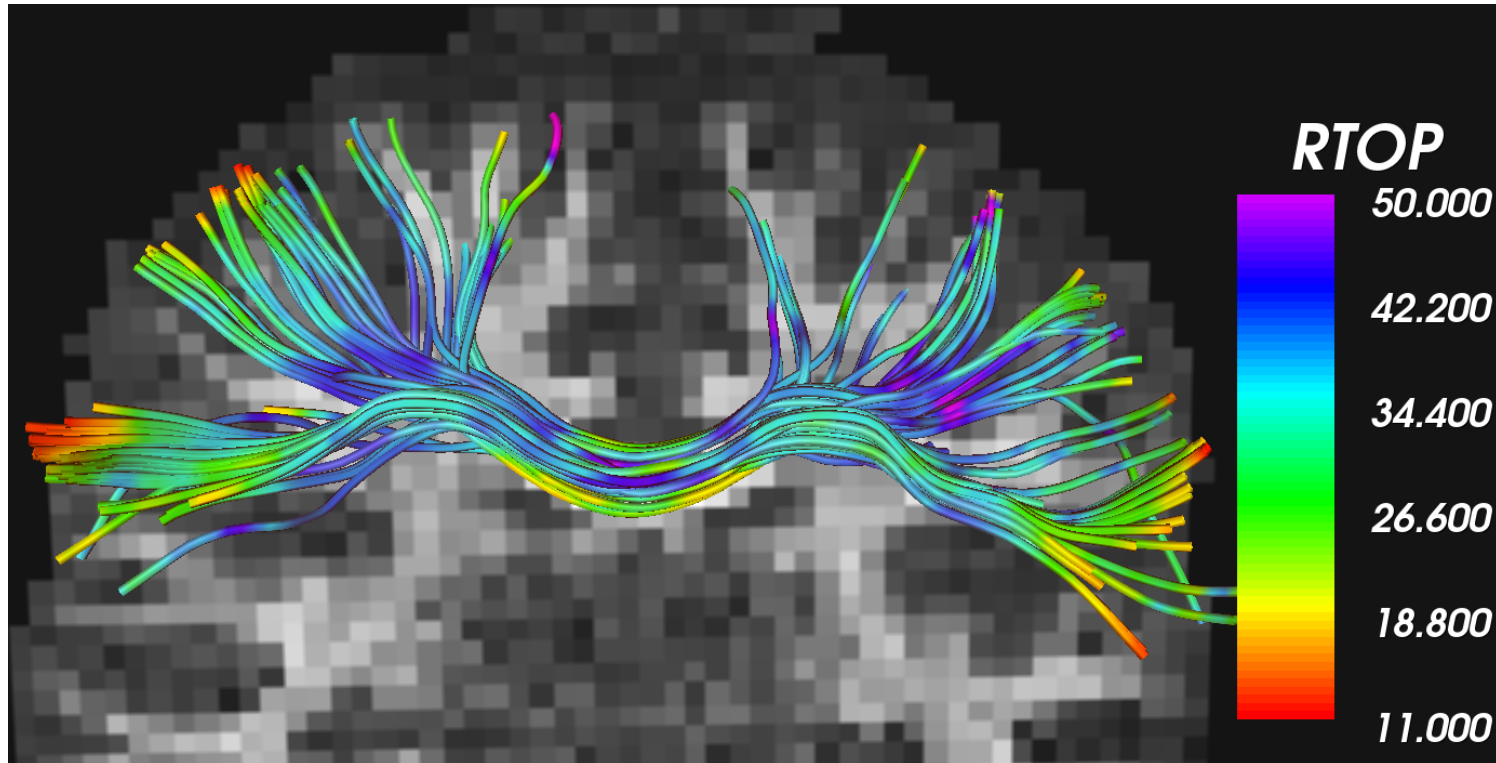
RTOP – return-to-origin probability – is the probability that a water molecule returns to its starting position (computed from the diffusion propagator).

Sparse reconstruction with UKF

Precentral inter-hemispheric fibers obtained with different number of gradient directions : Red – 60/shell, Green – 30/shell, Blue – 20/shell and white – 16/shell. We used 2 shells with a b-value of 1000 and 4000.
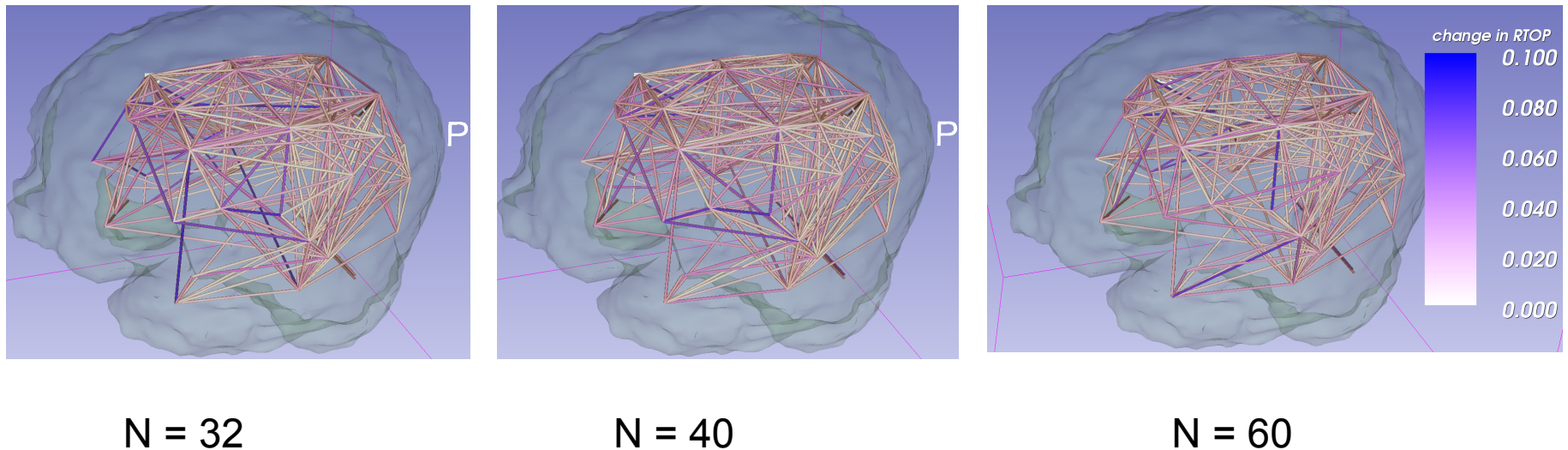
Sparse reconstruction with UKF

N = 32   N = 40   N = 60
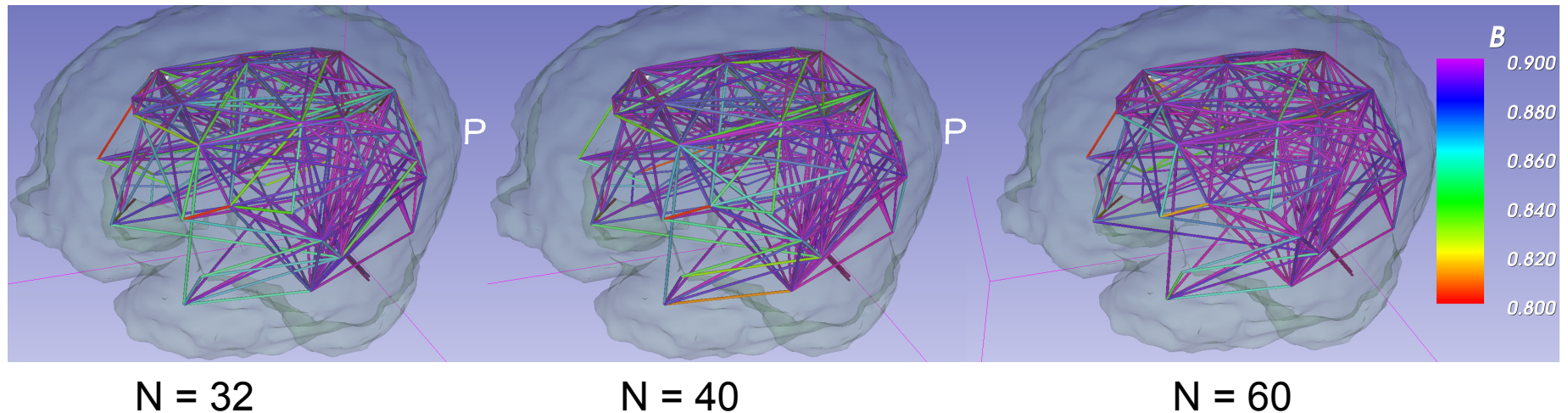
Fiber bundle overlap between the gold standard of N=120
and those obtained with fewer gradient directions.

<u>Inference:</u> Most fiber bundles, even with N=32, show a very good overlap of
around 0.9 and all fiber bundles have overlap above 0.8. This implies, very
good directional information is preserved even with fewer measurements.

MICCAI 2013.

Sparse reconstruction with UKF

N = 32   N = 40   N = 60

RTOP estimation difference between the gold standard of N=120 and that estimated with fewer gradient directions.

Inference: Most fiber bundles, even with N=32, show a very good reproducibility of RTOP with the difference in estimation similar to that obtained in test-retest DTI studies of FA.

MICCAI 2013.

Conclusion

68