# *Forward Models and Bayesian Inference in Diffuse Optical Brain Imaging*

May 29, 2003 at Mount Sinai School of Medicine

Alex Barnett Courant Institute, NYU

Collaborators (*NMR Center, Mass. Gen. Hosp., Boston*)

David Boas Joe Culver Anna Custo (MIT)

Gregory Sorensen Anders Dale

Funding: NIH, CIMIT

# Big picture

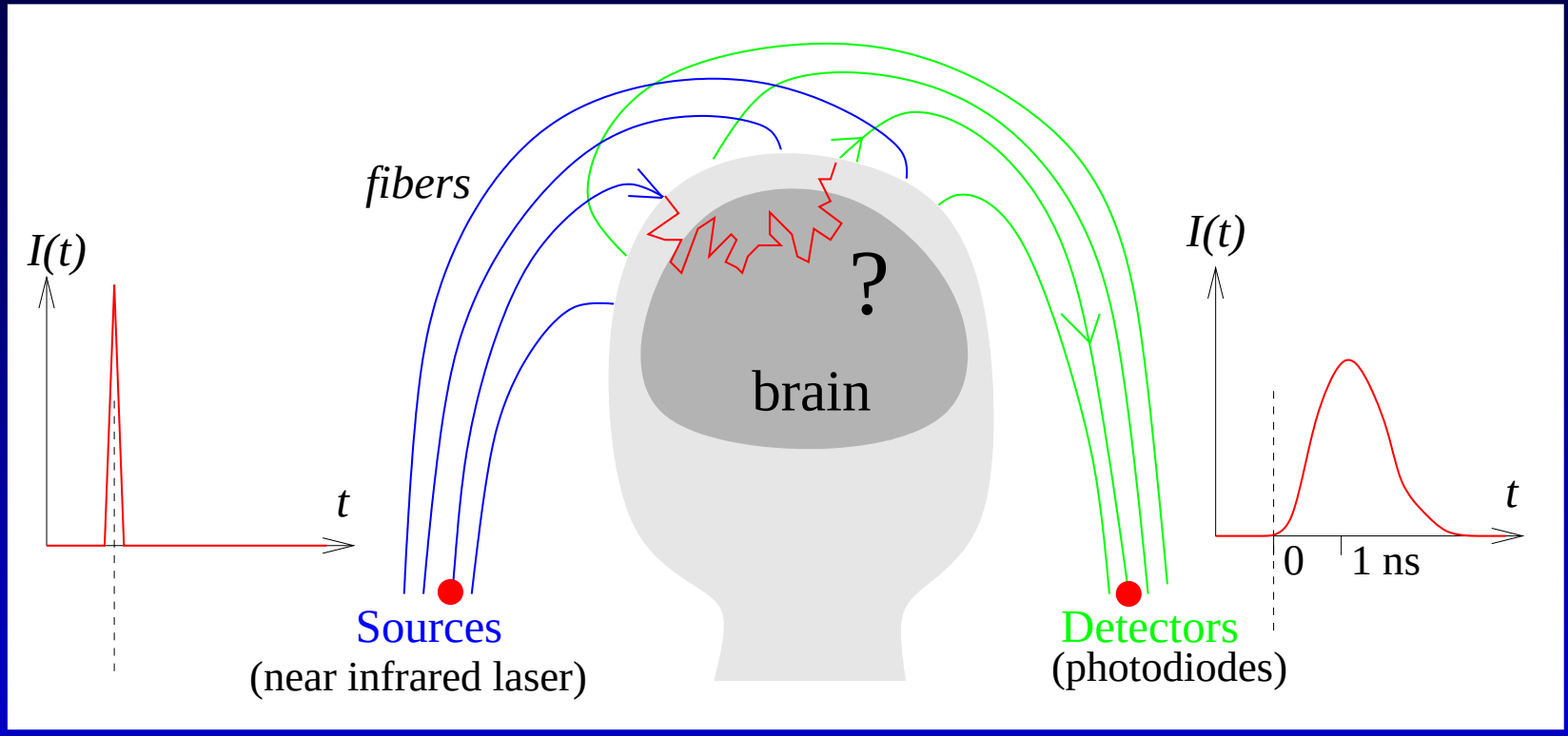
What can access optically, from outside the body?

DOT : ‘Diffuse Optical Tomography’

# Big picture

What can access optically, from outside the body?

DOT : ‘Diffuse Optical Tomography’




Tissue is highly scattering (blurring)

Get spatial maps { absorption / scattering } at some wavelength(s)?

# Outline

- motivation
- background & history
- physics of light in tissue
- problem: MRI-segmented geometry
  - measure *baseline* optical properties
- forward models
- inverse problem via Bayes
- some results
- Markov chain sampling, calibration…
- conclusions

# Brain: what interested in?

CLINICAL

- Cerebral oximetry (*e.g.* neonatal):
  - absolute oximetry hard
- Imaging stroke (local lack of $O_2$), hemorrhage

# Brain: what interested in?

CLINICAL

- Cerebral oximetry (*e.g.* neonatal):
  - absolute oximetry hard
- Imaging stroke (local lack of $O_2$), hemorrhage

FUNCTIONAL

- Organization of brain:
  - neural response as func of space & time
- functional disorders: seizure, depression

# Brain: what interested in?

CLINICAL

- Cerebral oximetry (*e.g.* neonatal):
  - absolute oximetry hard
- Imaging stroke (local lack of $O_2$), hemorrhage

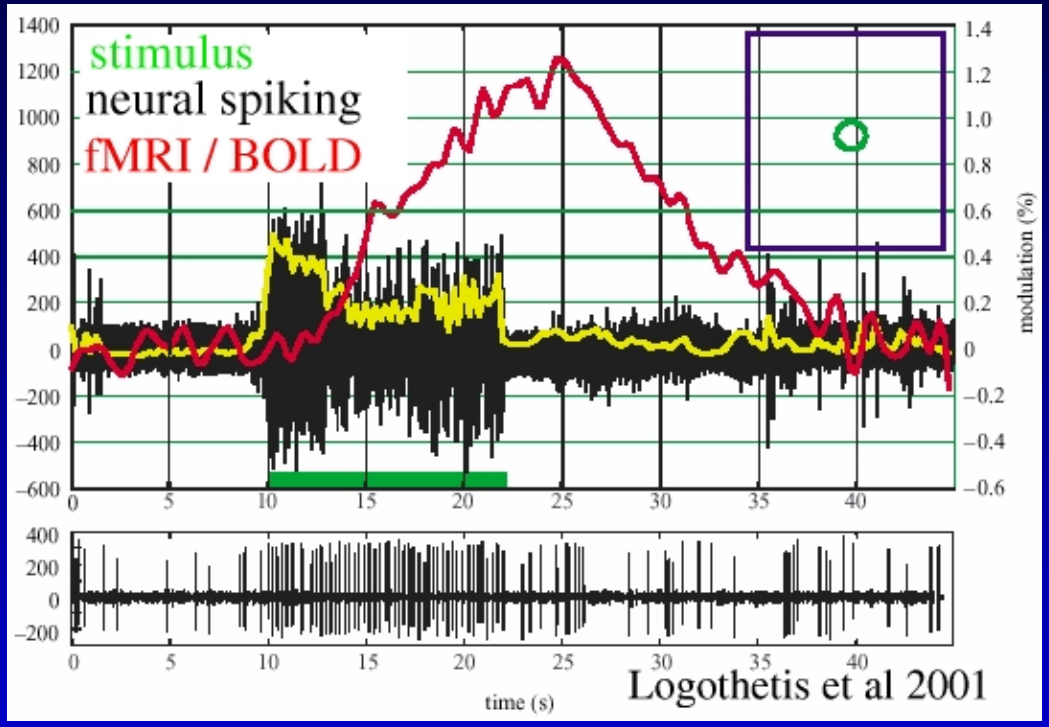
FUNCTIONAL

- Organization of brain:
  - neural response as func of space & time
- functional disorders: seizure, depression

[also breast tumors, arthritis, muscle oximetry. . . ]

# Functional imaging: why blood?

Detect neural firing { microelectrodes — ouch!
MEG — costly, low resolution




'Hemodynamic Response Function'

Neural activation → increased blood flow

*1990s*: functional MRI revolution (brain mapping)

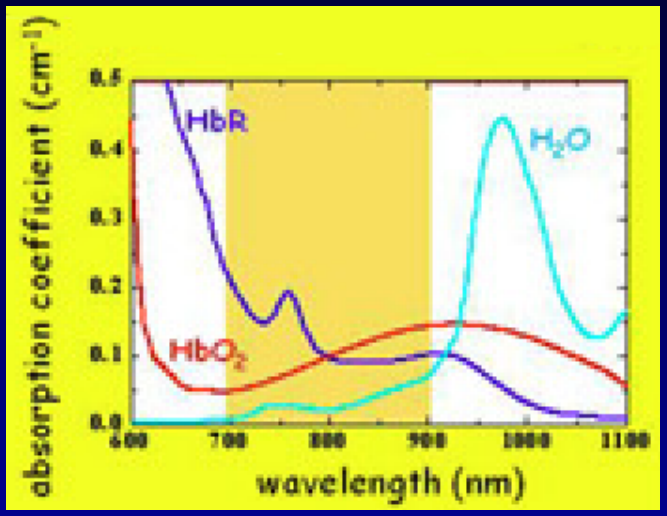
# Optical spectroscopy of the body

Near IR ‘window’ 700-900 nm :

- absorption $\mu_a$ is low
- hemoglobin dominates $\mu_a$

HbR - deoxy

$HbO_2$ - oxy



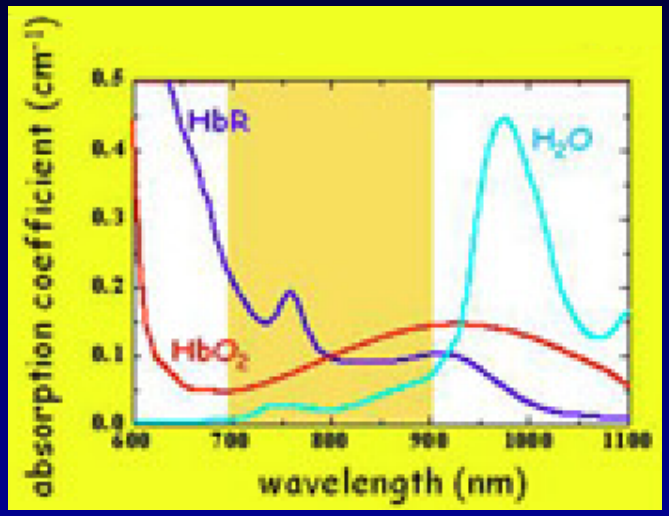
# Optical spectroscopy of the body

Near IR ‘window’ 700-900 nm :

- absorption $\mu_a$ is low
- hemoglobin dominates $\mu_a$

HbR - deoxy

$HbO_2$ - oxy




*von Vierordt 1876*: spectroscope, light through finger

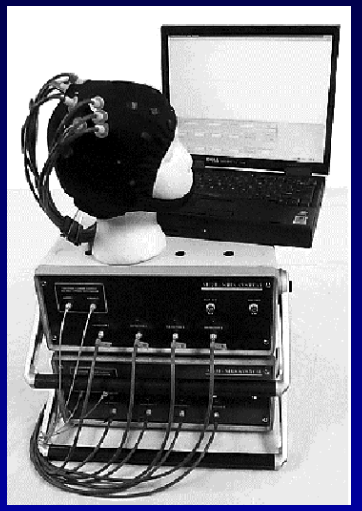
*Millikan 1940*: wartime fighter pilot oximeter

*Aoyagi 1970s*: pulse oximetry $\longrightarrow$ clinical

*Jobsis 1980s*: first noninvasive brain activation

*1990s*: functional brain mapping

# Current DOT apparatus



State of the art. . .

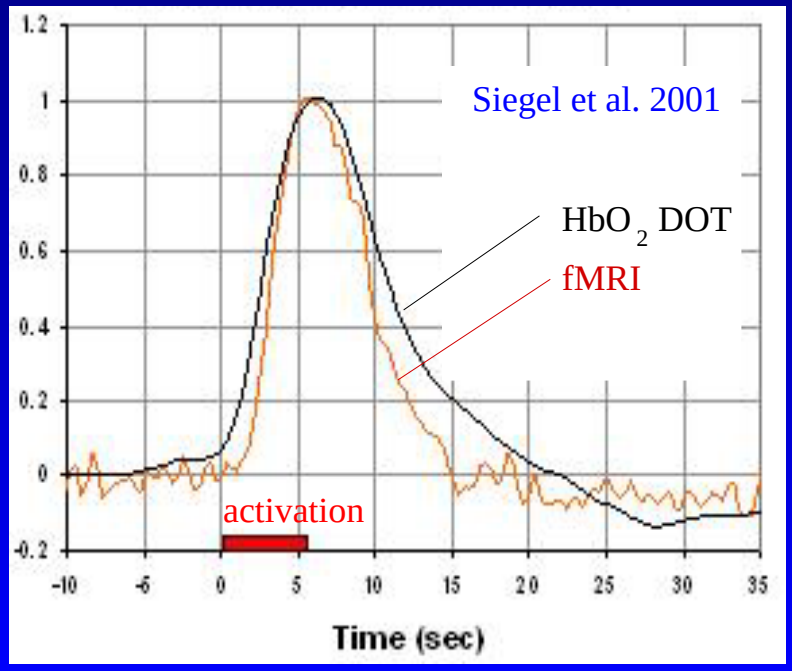
- up to typ. 32 source $\times$ 32 detector
- single photon counting (time-correlated)
- time-resolution of 10 ps (but blurred by $\sim$ 100 ps)
- several simultaneous wavelengths

# Compare DOT vs fMRI

| | fMRI | DOT |
|---|---|---|
| space | 2–4 mm | 1–2 cm, not too deep |
| time | 1–2 s | 10–100 ms |
| portable | no | yes (head can move) |
| cost | $> \$10^6$ | $\leq \$10^5$ |
| sens | HbR only | both $HbO_2$ and HbR |

# Compare DOT vs fMRI

| | fMRI | DOT |
|---|---|---|
| space | 2–4 mm | 1–2 cm, not too deep |
| time | 1–2 s | 10–100 ms |
| portable | no | yes (head can move) |
| cost | $> \$10^6$ | $\leq \$10^5$ |
| sens | HbR only | both $HbO_2$ and HbR |




Ongoing:

- validate DOT vs fMRI
- model: neural $\overset{?}{\leftrightarrow}$ vascular
- what does fMRI measure?

# Photon migration

Incoherent light $\longrightarrow$ ballistic transport of $f(\mathbf{r}, \mathbf{\Omega}, t)$:

$$\begin{aligned} \tfrac{1}{c}\dot{f} \;=\; & -\,\mathbf{\Omega}\cdot\nabla f && \text{flow} \\ & -\,[\mu_a(\mathbf{r}) + \mu_s(\mathbf{r})]f && \text{leaving} \\ & +\,\int d\mathbf{\Omega}'\, S(\mathbf{r};\mathbf{\Omega},\mathbf{\Omega}')\, f(\mathbf{r},\mathbf{\Omega}') && \text{arriving} \\ & +\,Q(\mathbf{r},\mathbf{\Omega}) && \text{source} \end{aligned}$$

speed $c$, absorption $\mu_a$, scattering $\mu_s = \int d\mathbf{\Omega}\, S$

# Photon migration

Incoherent light $\rightarrow$ ballistic transport of $f(\mathbf{r}, \mathbf{\Omega}, t)$:

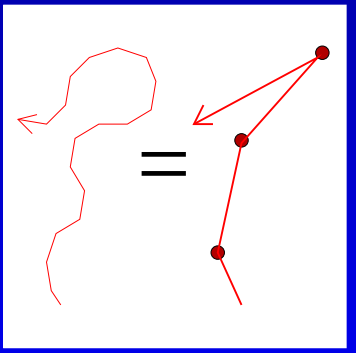
$$\begin{aligned} \tfrac{1}{c}\dot{f} \;=\; & -\,\mathbf{\Omega}\cdot\nabla f && \text{flow} \\ & -\,[\mu_a(\mathbf{r}) + \mu_s(\mathbf{r})]f && \text{leaving} \\ & +\,\textstyle\int d\mathbf{\Omega}'\, S(\mathbf{r};\mathbf{\Omega},\mathbf{\Omega}')\, f(\mathbf{r},\mathbf{\Omega}') && \text{arriving} \\ & +\,Q(\mathbf{r},\mathbf{\Omega}) && \text{source} \end{aligned}$$

speed $c$, absorption $\mu_a$, scattering $\mu_s = \int d\mathbf{\Omega}\, S$

- Legendre $f(\mathbf{r}, \mathbf{\Omega}) = \phi(\mathbf{r}) + \mathbf{j}(\mathbf{r}) \cdot \mathbf{\Omega} +$ ignored $\ldots$

$\mathbf{j}$ small, relaxes (fast) to be $\propto \nabla\phi$
$\phi$ diffuses (slow), with coeff $1/3\mu_s'$

'reduced scatt' $\mu_s' = (1 - \langle\cos\theta\rangle_S)\mu_s$



# Diffusion approximation

you need: $\mu_a \ll \mu_s'$, length scales $\gg \frac{1}{\mu_s'}$

$$\frac{1}{c}\dot{\phi} = \nabla \cdot \left( \frac{1}{3\mu_s'(\mathbf{r})} \nabla \phi \right) - \mu_a(\mathbf{r})\phi + q(\mathbf{r}, t)$$

$\phi =$ fluence  Time-evolution of parabolic PDE

# Diffusion approximation

you need: $\mu_a \ll \mu_s'$, length scales $\gg \frac{1}{\mu_s'}$

$$\frac{1}{c}\dot{\phi} = \nabla \cdot \left( \frac{1}{3\mu_s'(\mathbf{r})} \nabla \phi \right) - \mu_a(\mathbf{r})\phi + q(\mathbf{r}, t)$$

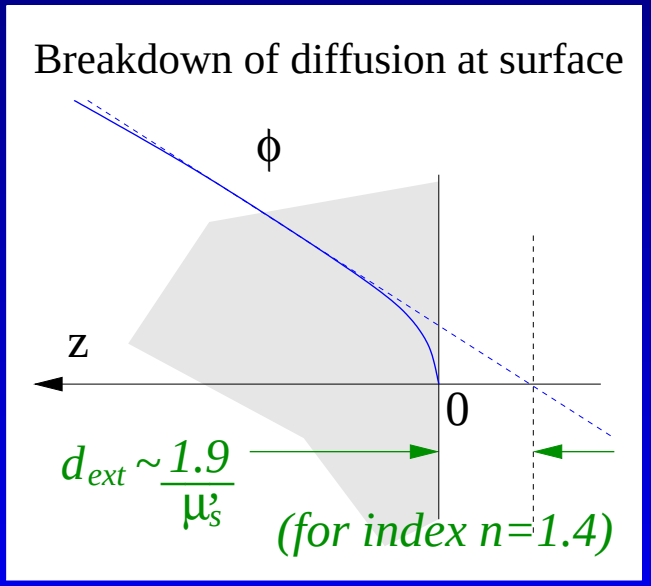
$\phi =$ fluence  Time-evolution of parabolic PDE

Asymptotic Dirichlet BC:

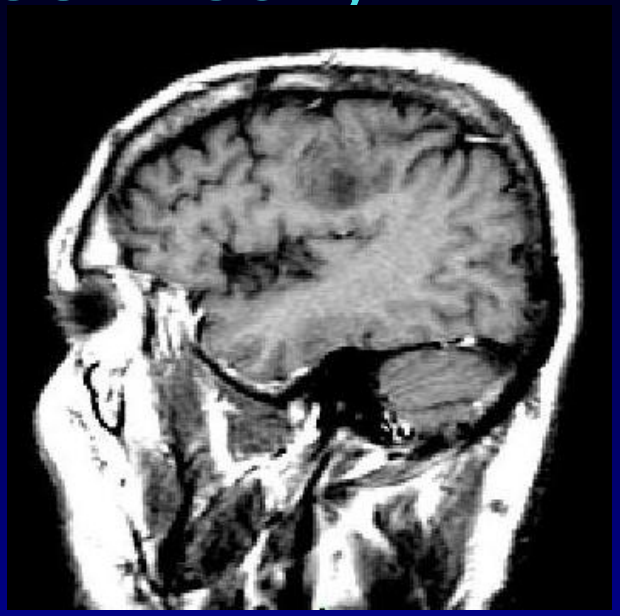
$$\phi\big|_{z=-d_{\text{ext}}} = 0$$

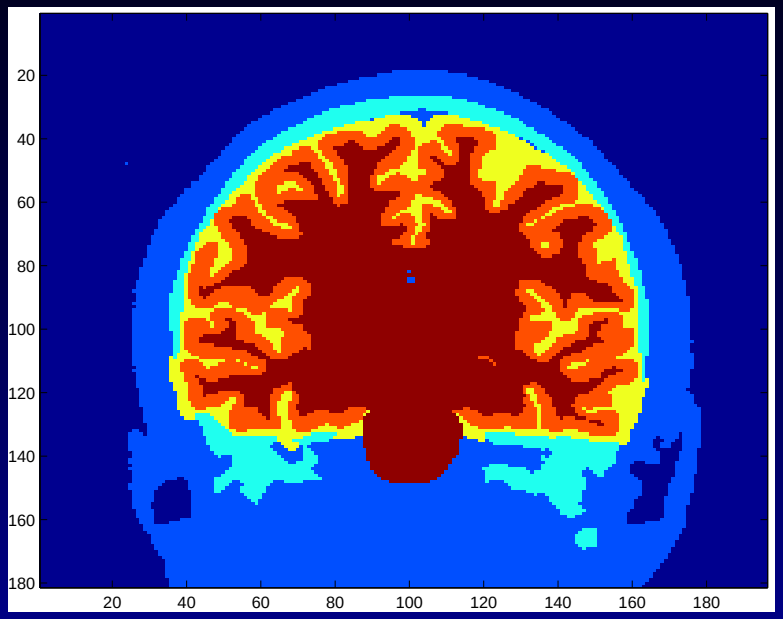
Source: $q = (\text{point src}) \cdot \delta(t)$ at $z = -1/\mu_s'$

Detector: meas. $\alpha\phi + \beta\partial\phi/\partial z$ (people argue $\alpha, \beta$)



# Geometry





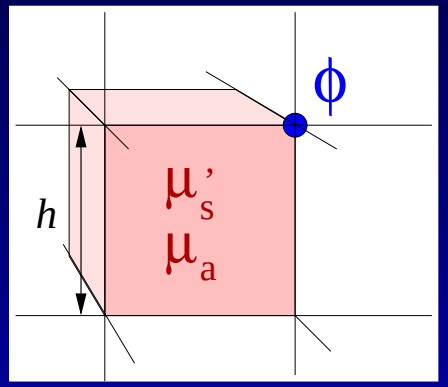

| tissue | $\mu_a$ (mm$^{-1}$) | $\mu_s'$ (mm$^{-1}$) | shape |
|---|---|---|---|
| scalp | 0.015 | 0.8 | $\sim$ 7 mm layer |
| skull | 0.01 | 1.0 | $\sim$ 7 mm layer |
| CSF | 0.0004 | 0.01* | folded 1–3 mm sheet |
| brain | 0.018 | 1.3 | $\sim$ 1 cm folds (sulci) |

Much uncertainty. *handled specially: $\mu_{s,\text{eff}}' \sim 0.4$ mm$^{-1}$.

# Forward model I: 3D segmented

Finite difference lattice, arbitrary tissue geom

- Forward Euler timestepping $\Delta t$
- accuracy $O(h^2)$, $O(\Delta t)$
- slow: effort $O(h^{-5})$
- $h = 2$ mm takes 8 sec CPU



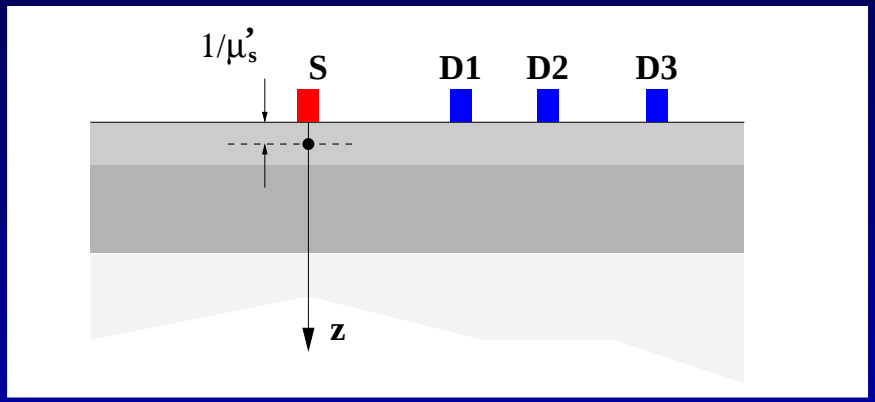

*Issues:*

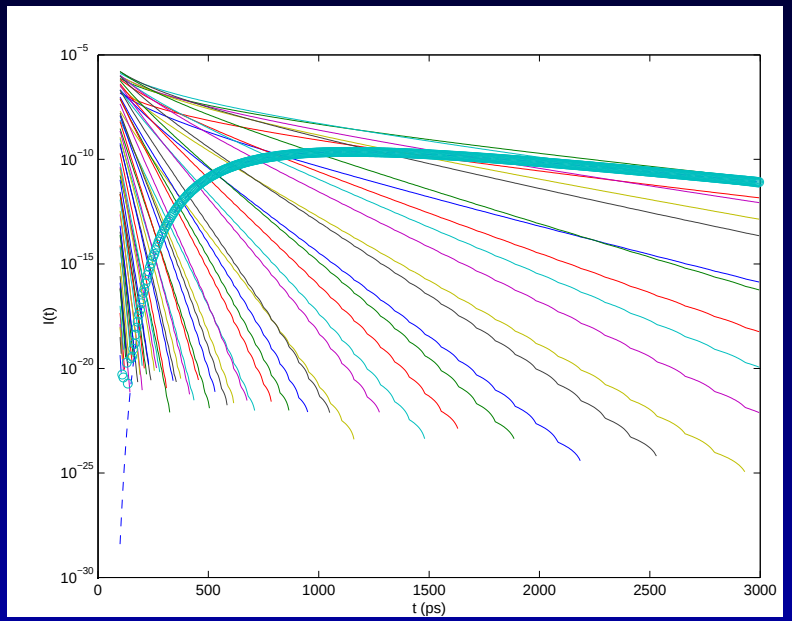
- Crank-Nicholson timestepping common in DOT.
  dangerous: not L-stable!
- Scalp surface voxellated $\rightarrow$ bad errors, BCs
  smooth BC would be much better...

# Forward model II: layered slab

Cyl. symm $\Rightarrow \phi(\mathbf{r}, t) = \sum$ transverse Bessel modes

Make use of $t \geq t_{\text{min}}$:
higher modes have died.



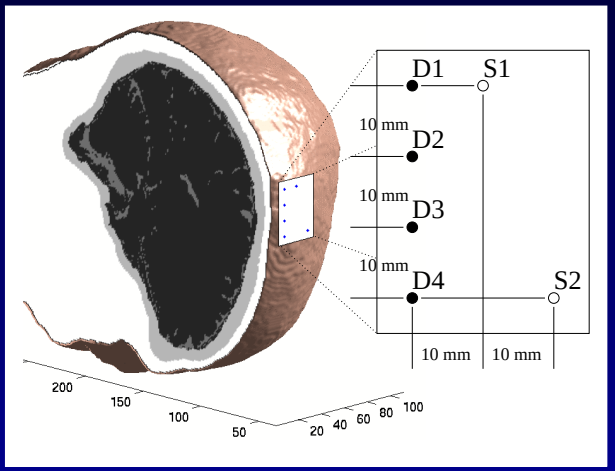

- Each mode is a 1D finite difference problem.
- novel Crank-Nicholson timestep $\Delta t = \beta t$,
  kills nonphysical modes, effort $O(\log \frac{t_{\text{max}}}{t_{\text{min}}})$
- Careful analysis: 1% error for *relevant* signals.
- Fast! typ. 0.2 sec CPU.

# Segmented ‘imaging’

Q: how well could we measure homogeneous tissues?

$\mathbf{x} \equiv \{\mu_a, \mu_s'\}$ scalp, skull, brain

- meas *absolute* values is hard
- needed for *quantitative* imaging of brain activation

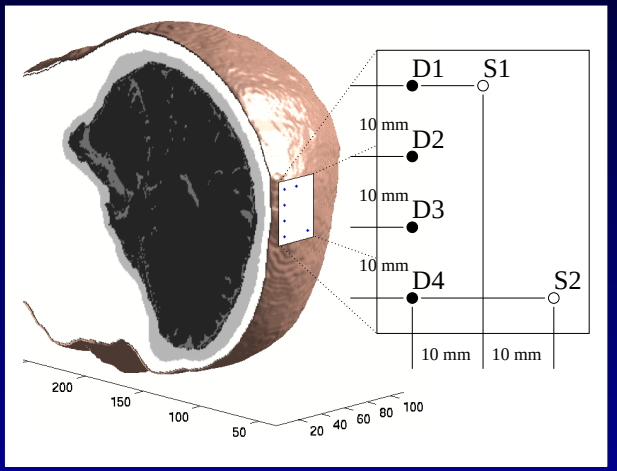


# Segmented ‘imaging’

Q: how well could we measure homogeneous tissues?

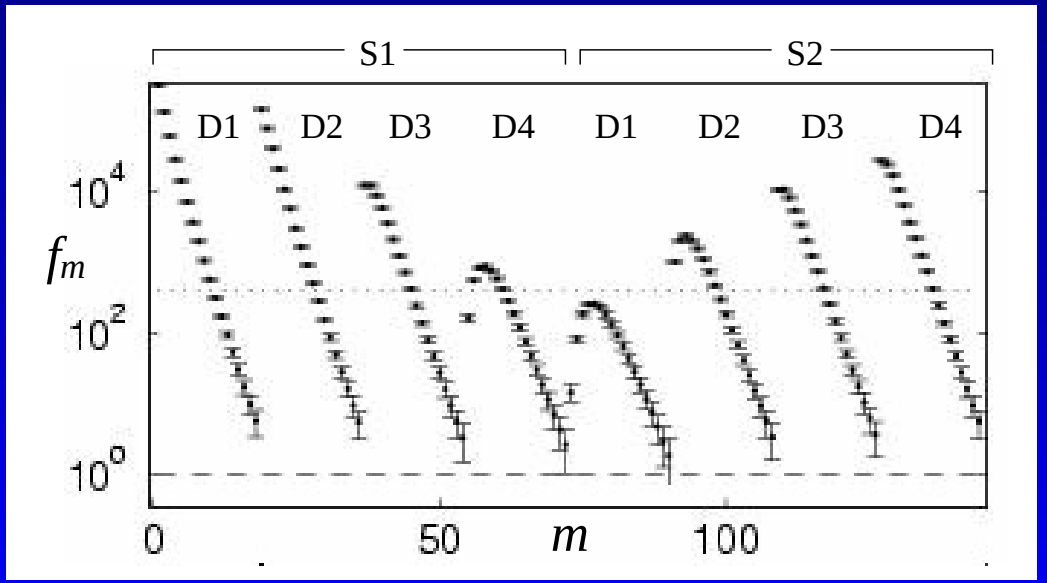
$\mathbf{x} \equiv \{\mu_a, \mu_s'\}$ scalp, skull, brain

- meas *absolute* values is hard
- needed for *quantitative* imaging of brain activation




SIMULATED SIGNAL $= \mathbf{y}$

Small DOT system (2 Src, 4 Det)



# Inverse problem

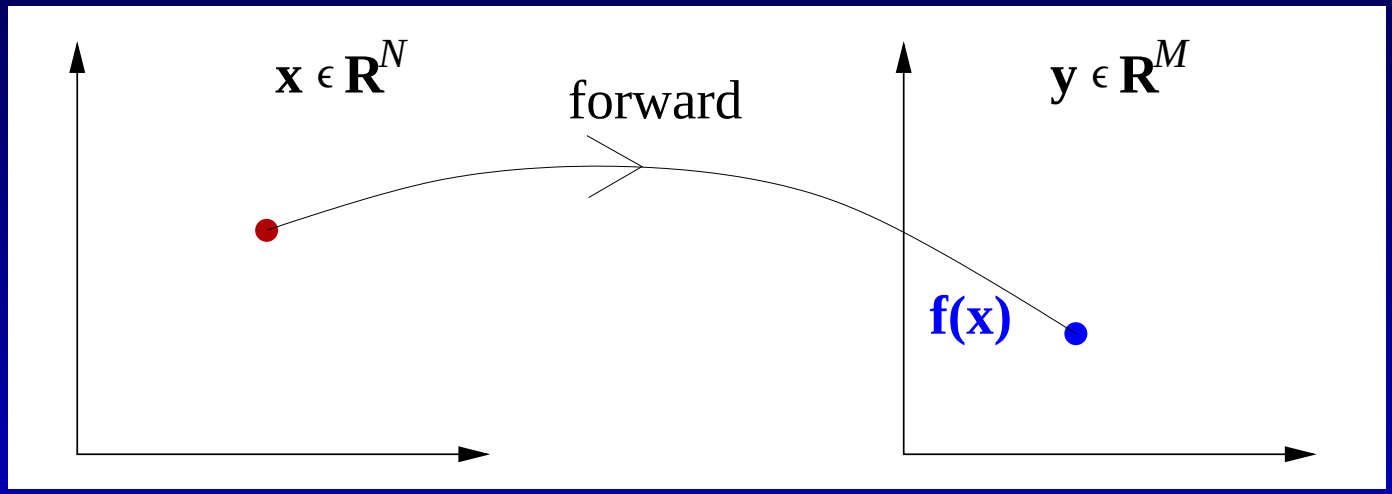
$$\mathbf{x} \equiv \{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\} \quad \overset{\text{forward}}{\longrightarrow} \quad \mathbf{f}(\mathbf{x})$$

parameter vector — expected signal vector

# Inverse problem

$$\mathbf{x} \equiv \{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\} \quad \xrightarrow{\text{forward}} \quad \mathbf{f}(\mathbf{x})$$

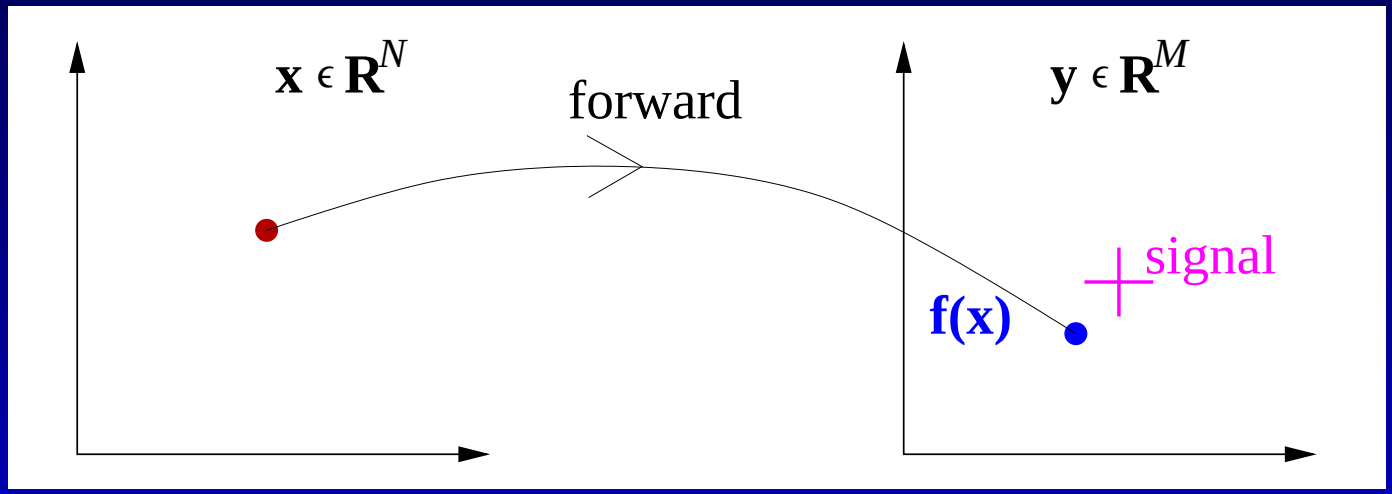
parameter vector — expected signal vector



# Inverse problem

$$\mathbf{x} \equiv \{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\} \quad \overset{\text{forward}}{\longrightarrow} \quad \mathbf{f}(\mathbf{x})$$

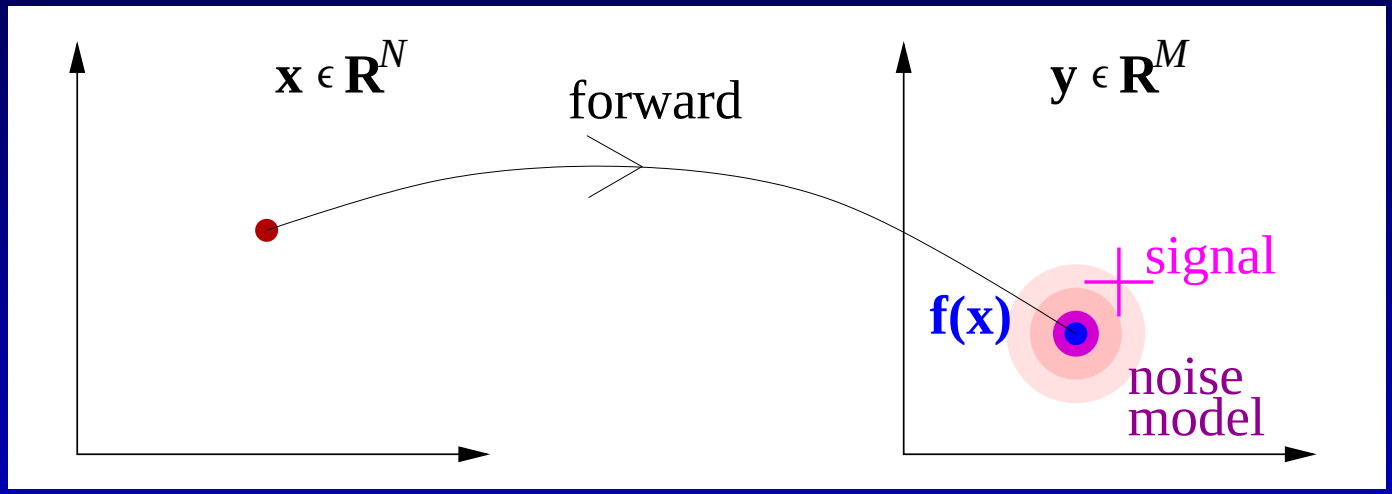
parameter vector — expected signal vector



# Inverse problem

$$\mathbf{x} \equiv \{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\} \xrightarrow{\text{forward}} \mathbf{f}(\mathbf{x})$$

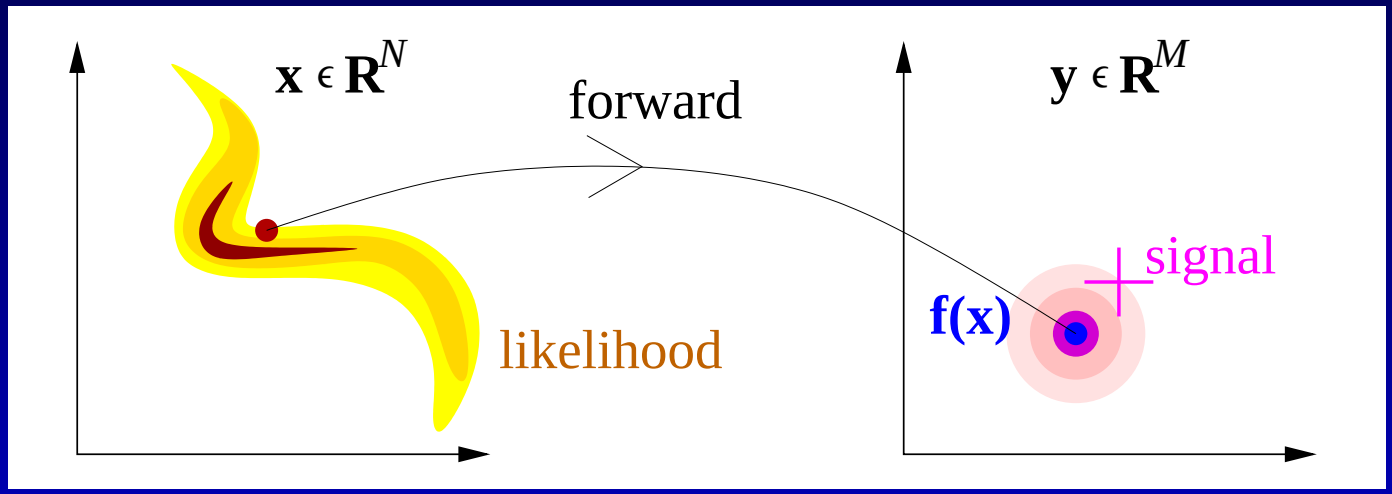
parameter vector — expected signal vector



# Inverse problem

$$\mathbf{x} \equiv \{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\} \quad \overset{\text{forward}}{\longrightarrow} \quad \mathbf{f}(\mathbf{x})$$

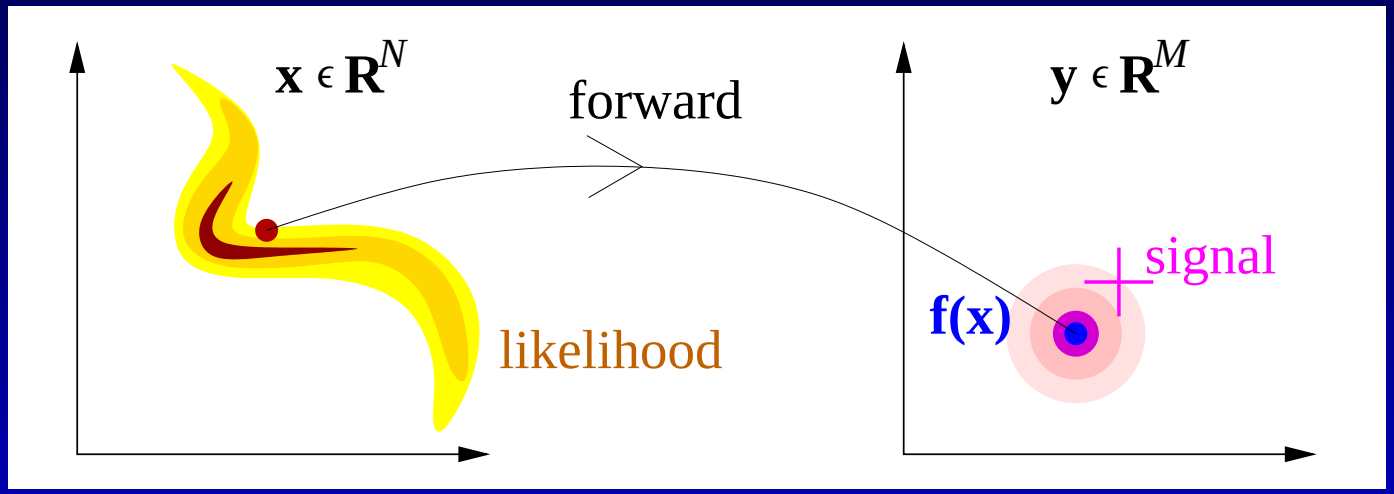
parameter vector — expected signal vector



# Inverse problem

$$\mathbf{x} \equiv \{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\} \quad \overset{\text{forward}}{\longrightarrow} \quad \mathbf{f}(\mathbf{x})$$

parameter vector — expected signal vector




Nonlinear $N$-dim optimization problem

$\frac{\partial f_m}{\partial x_n}$ sing. vals. $\rightarrow 0$ : 'ill-posed' (many x equally valid)

# Bayesian statistical method

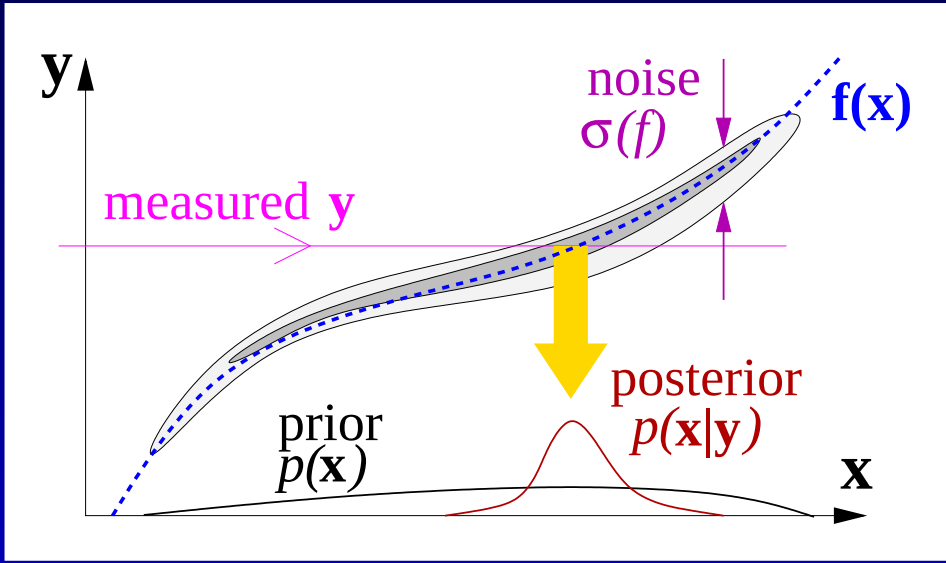
Incomplete info on $\mathbf{x} \rightarrow$ *probability density function*

Entire model = joint PDF $p(\mathbf{x}, \mathbf{y})$

# Bayesian statistical method

Incomplete info on $\mathbf{x} \rightarrow$ *probability density function*

Entire model = joint PDF $p(\mathbf{x}, \mathbf{y})$




Bayesian inference

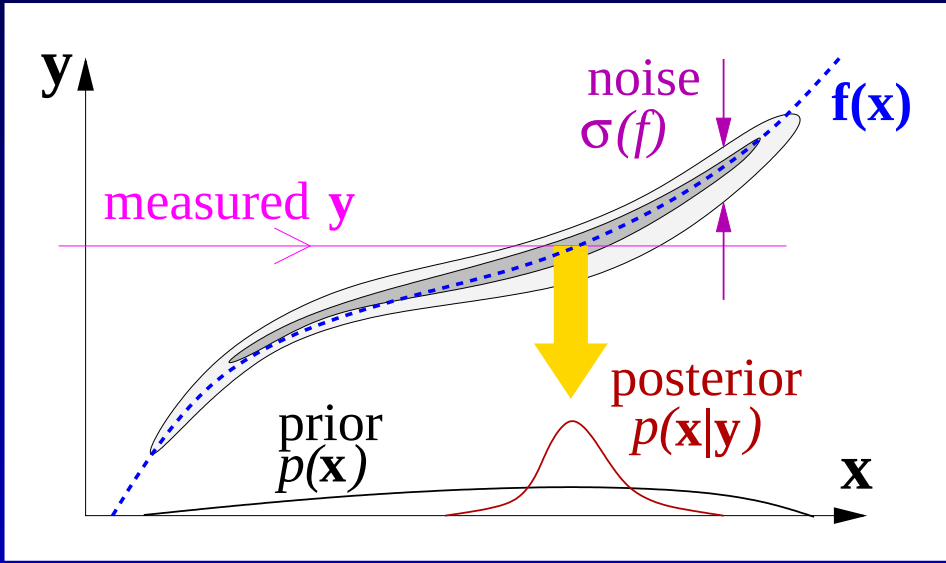
$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x}, \mathbf{y})$$
$$= p(\mathbf{y}|\mathbf{x}) \cdot p(\mathbf{x})$$

posterior likelihood prior

# Bayesian statistical method

Incomplete info on $\mathbf{x}$ $\longrightarrow$ *probability density function*

Entire model = joint PDF $p(\mathbf{x}, \mathbf{y})$




Bayesian inference

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x}, \mathbf{y})$$
$$= p(\mathbf{y}|\mathbf{x}) \cdot p(\mathbf{x})$$

posterior likelihood prior

- Embraces ill-posedness, no 'overfitting'.
- Makes assumptions (*e.g.* noise model) explicit.
- Need to explore $N$-dim posterior: many $\mathbf{f}(\mathbf{x})$ evals required ($> 10^2$)

# Why full PDF?

Usual ‘Bayesian’ method for DOT inverse problem: find single best-fit $\mathbf{x} = \mathbf{x}_{\mathrm{MAP}}$. (MAP = *maximum a-posteriori*)

# Why full PDF?

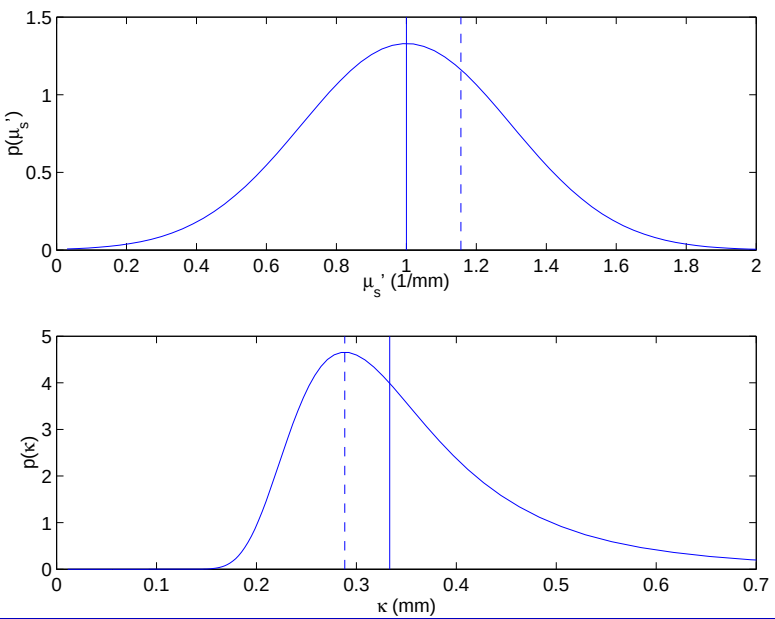
Usual ‘Bayesian’ method for DOT inverse problem: find single best-fit $\mathbf{x} = \mathbf{x}_{\mathrm{MAP}}$. (MAP = *maximum a-posteriori*)

However $\mathbf{x}_{\mathrm{MAP}}$ can be moved by reparametrization!

# Why full PDF?

Usual ‘Bayesian’ method for DOT inverse problem: find single best-fit $\mathbf{x} = \mathbf{x}_{\mathrm{MAP}}$. (MAP = *maximum a-posteriori*)

However $\mathbf{x}_{\mathrm{MAP}}$ can be moved by reparametrization!




Example: broad PDF on $\mu_s'$
Normal, $\sigma = 30\%$ of mean.

16% shift in MAP between $p(\mu_s')$ and $p(\mu_s'^{-1})$

- CPUs advance faster than DOT instrumentation
  $\rightarrow$ make maximal use of data
- Statistical answers $\rightarrow$ multimodal imaging.

# Realistic new noise model

Each signal component $m = 1 \cdots M$ independent.

Photons Poissonian: gaussian approx $\sigma(f) = f^{1/2}$

*E.g.* $10^6$ photons = 0.1% frac error

*But: we do not trust forward model to 0.1% !*

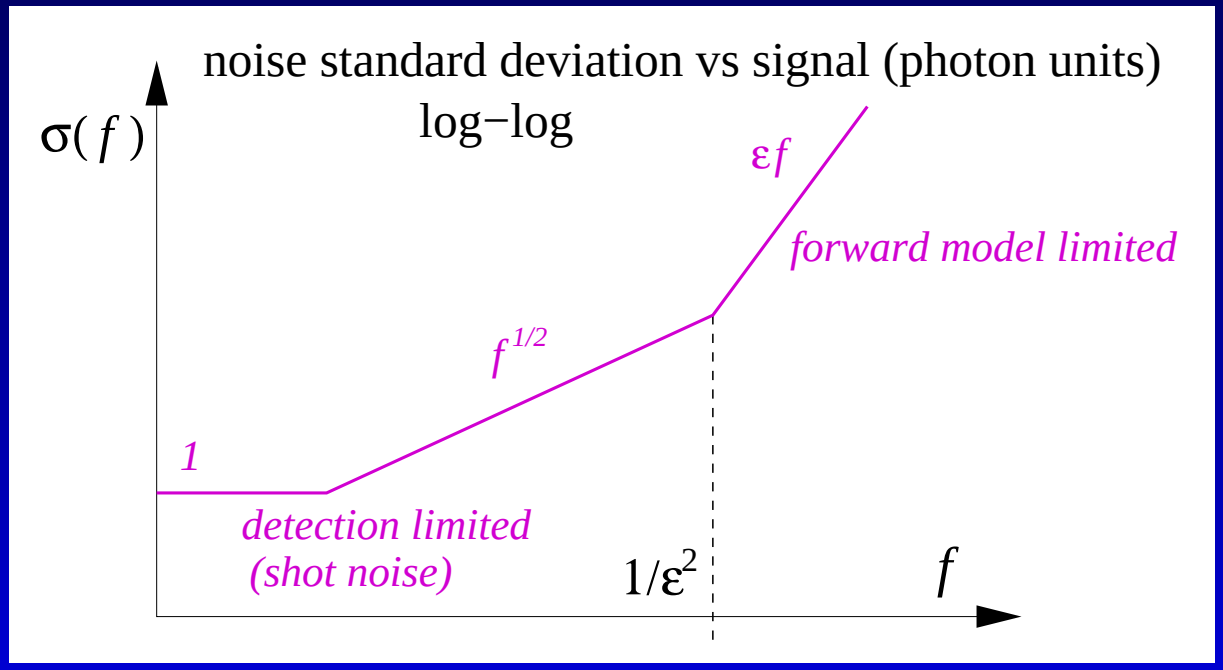
# Realistic new noise model

Each signal component $m = 1 \cdots M$ independent.

Photons Poissonian: gaussian approx $\sigma(f) = f^{1/2}$

*E.g.* $10^6$ photons = 0.1% frac error

*But: we do not trust forward model to 0.1% !*




- $\varepsilon =$ fractional forward model error *e.g.* 1–5% (errors: physical, numerical. . . )

# Posterior: Gaussian approx

Flat prior gives posterior $e^{-L(\mathbf{x}|\mathbf{y})}$, with

$$L(\mathbf{x}|\mathbf{y}) = \frac{1}{2}\sum_{m=1}^{M} \ln \sigma(f_m)^2 + \frac{1}{2}\sum_{m=1}^{M} \frac{(f_m(\mathbf{x}) - y_m)^2}{\sigma(f_m)^2}.$$

Approx: $L(\mathbf{x})$ quadratic about $\mathbf{x}_{\text{MAP}}$

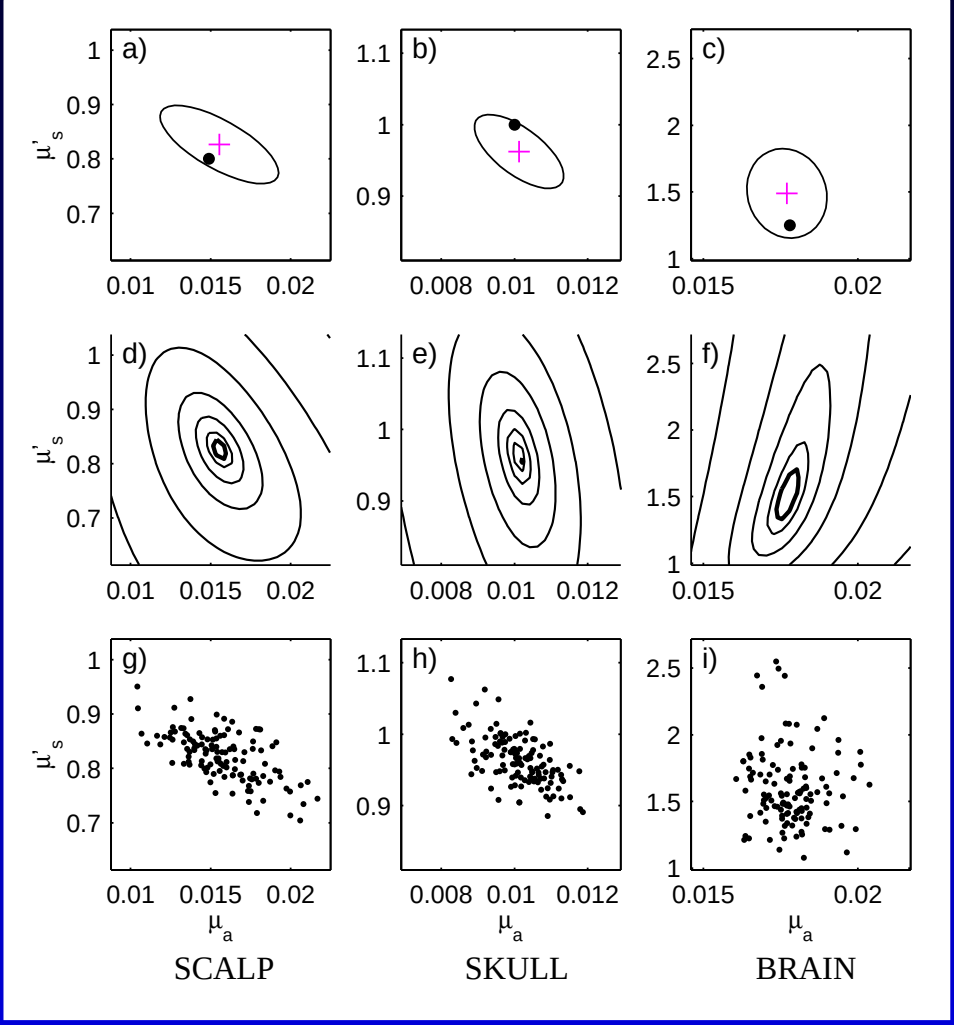
Our (Gaussian) noise model is lovable...

- $\nabla L$ exact given Jacobean $J_{mn} = \frac{\partial f_m}{\partial x_n}$.
- Good approx for Hessian $H = \nabla\nabla L$ is known.
- meas. $J$ via small changes in $\mathbf{x}$ (ok if $N$ small).

Optimization to find $\mathbf{x}_{\text{MAP}}$: trust region method (Levenburg-Marquardt), needs $\nabla L$ and $H$.

# Results: posterior PDF

Applied Optics, special issue on biomedical optics, June 2003.




Gaussian PDF approx

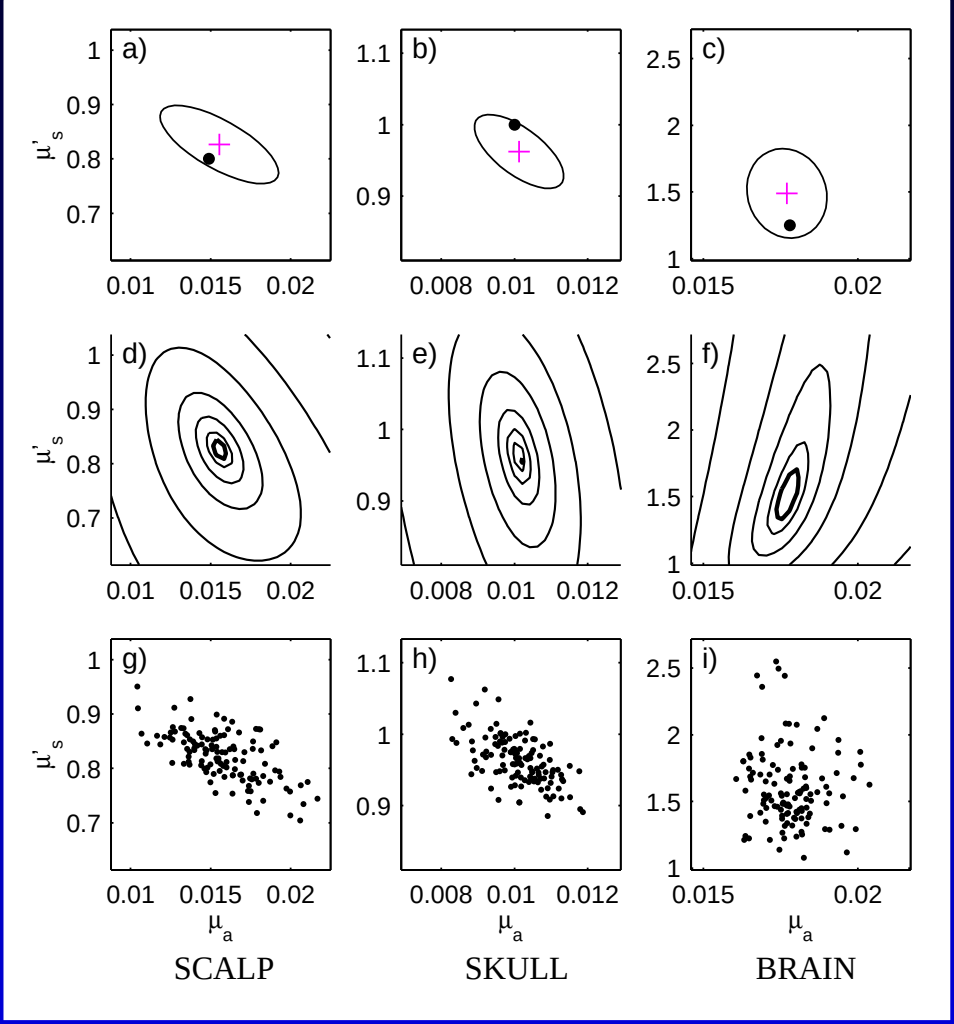
Marginal distributions, ellipse encloses 65% mass

*E.g.* $10^6$ det photons gives 5% in $\mu_a$, 20% in $\mu'_{s,\text{brain}}$

for $\varepsilon = 5\%$, flat prior

# Results: posterior PDF

Applied Optics, special issue on biomedical optics, June 2003.




Gaussian PDF approx

Marginal distributions, ellipse encloses 65% mass

Conditional (slices)

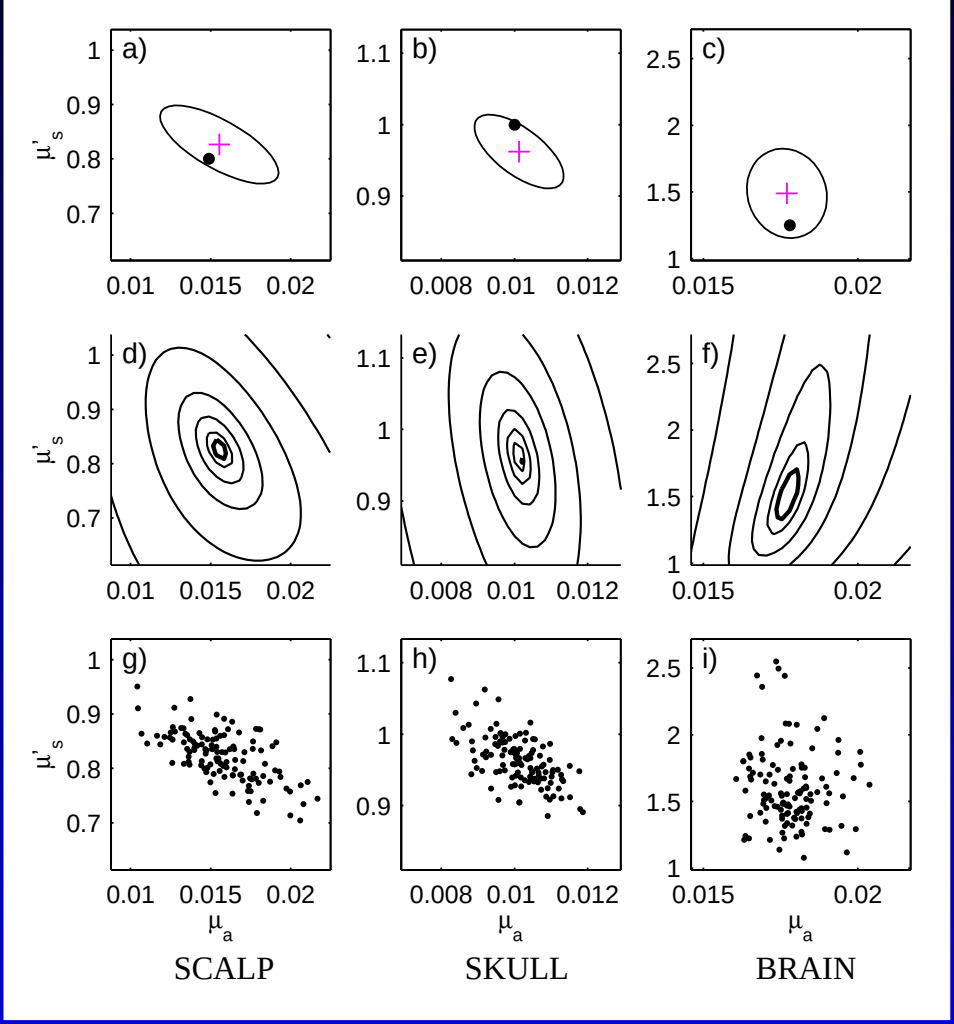
shows nonlinearity but not marginal shape

*E.g.* $10^6$ det photons gives 5% in $\mu_a$, 20% in $\mu'_{s,\text{brain}}$

for $\varepsilon = 5\%$, flat prior

# Results: posterior PDF

Applied Optics, special issue on biomedical optics, June 2003.




Gaussian PDF approx

Marginal distributions, ellipse encloses 65% mass

Conditional (slices)

shows nonlinearity but not marginal shape

Sampling exact PDF

marginal distributions via Markov chain Monte Carlo

*E.g.* $10^6$ det photons gives 5% in $\mu_a$, 20% in $\mu'_{s,\text{brain}}$

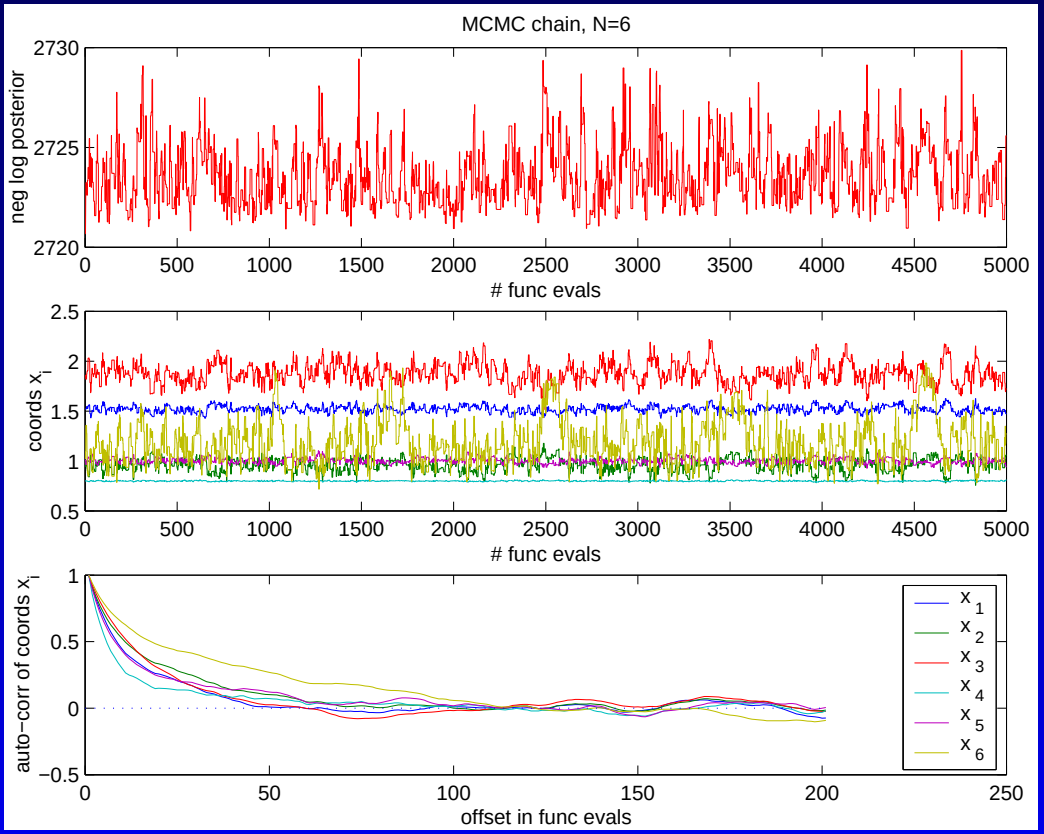
for $\varepsilon = 5\%$, flat prior

# Markov chain Monte Carlo

Want to sample $Q(\mathbf{x})$: transition rule $T$ obeys detailed balance,

$$T(\mathbf{x}' \leftarrow \mathbf{x})Q(\mathbf{x}) \;=\; T(\mathbf{x} \leftarrow \mathbf{x}')Q(\mathbf{x}').$$

Then $Q(\mathbf{x})$ is equilibrium dist. $T$ should 'mix' (explore $\mathbf{x}$) fast.

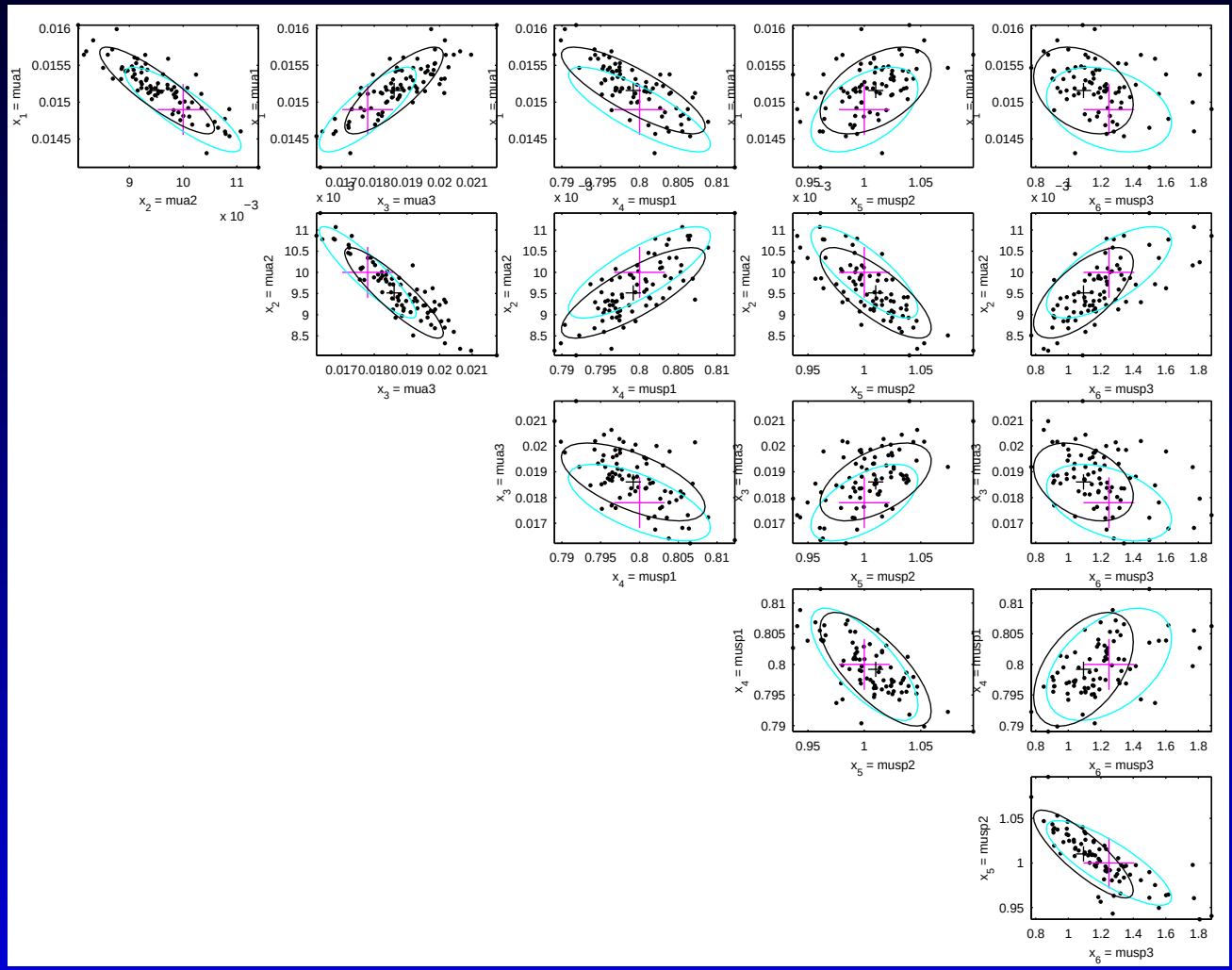



Crude $T$: Metropolis walk with steps $\Delta\mathbf{x}$, sampled from hyper-ellipsoid of Hessian $H_{\text{MAP}}$.

Get 1 indep. sample per corr length ($\sim 50$)

Faster methods exist

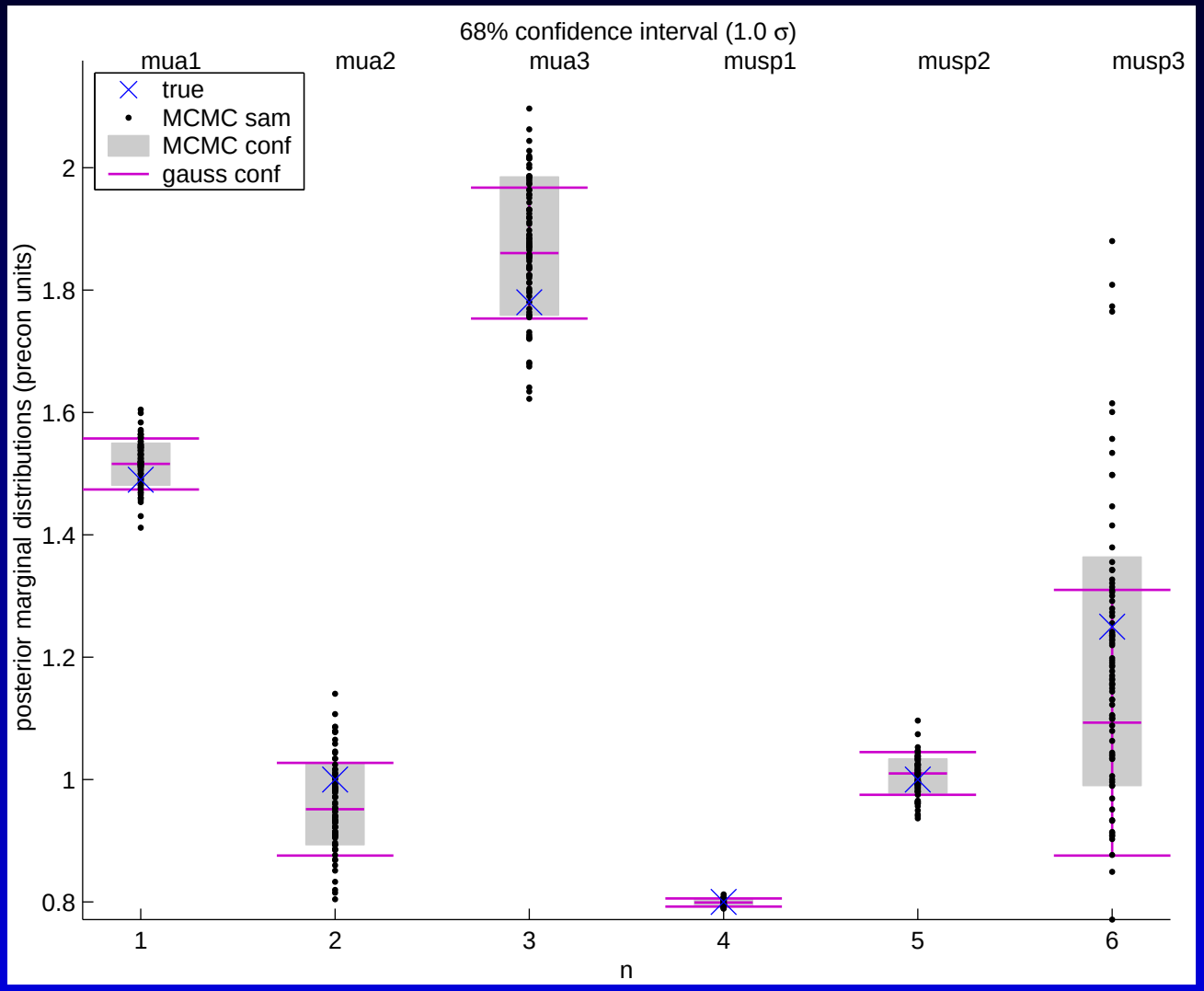
# Results: all 2d marginal PDFs




True value + falls within posterior cloud.

Pancake-like PDF: $\lambda_{\max}(H_{\mathrm{MAP}})/\lambda_{\min}(H_{\mathrm{MAP}}) \sim 10^4$

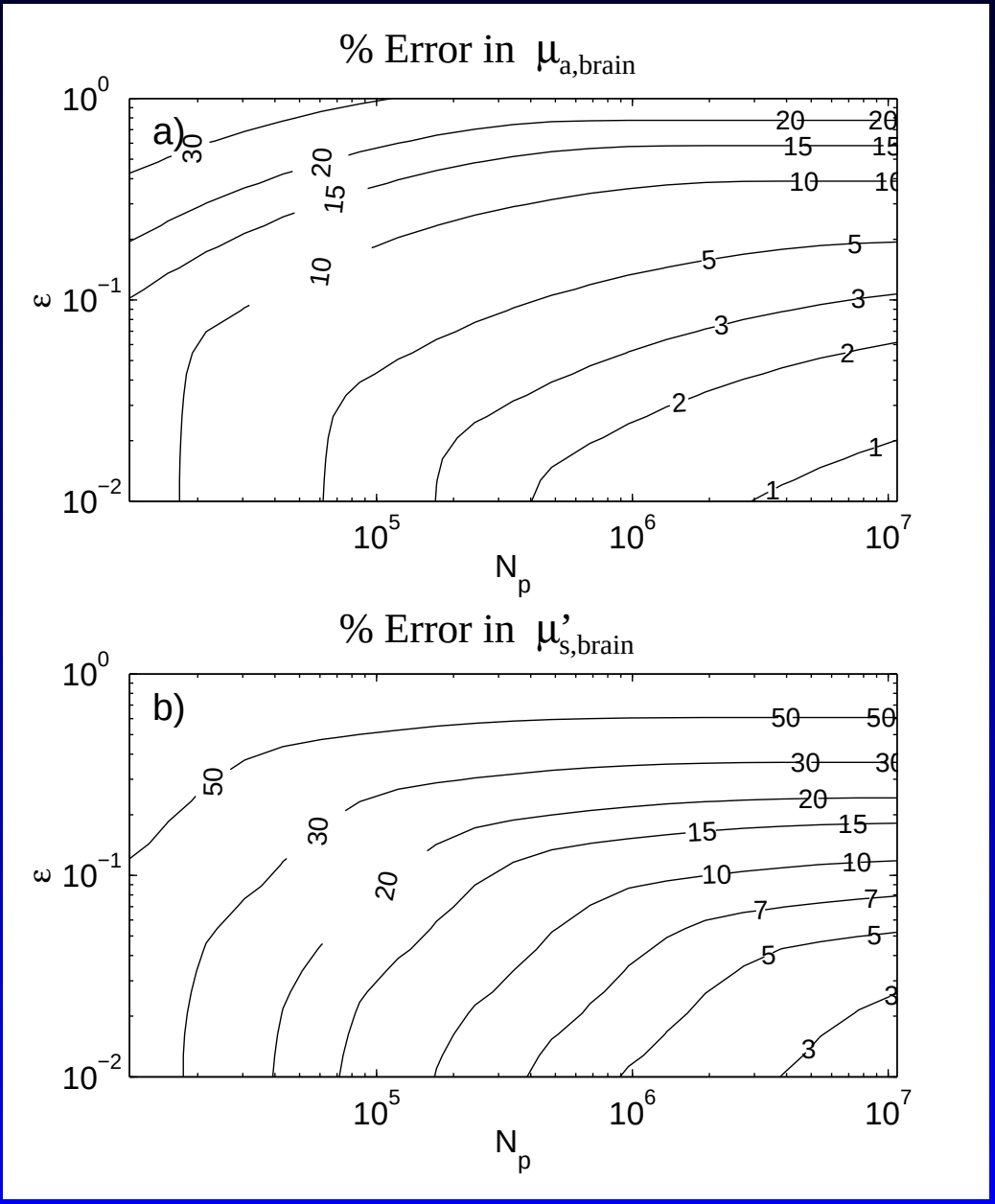
# Confidence intervals




Gaussian approx good apart from tails in $\mu'_{s,\text{brain}}$

[nonlinearity of $\mathbf{f}(\mathbf{x})$]

# How many photons needed?




optimize DOT apparatus design

$N_p$ = total detected photons

# Bayesian calibration

$\mathbf{c}$ = unknown calibration (nuisance) parameters.

Eg freckle/hair affects detector amplitude.

Want to fit them as part of inverse problem.

Standard DOT approach: find $\mathbf{c}_{\text{best fit}}$, then $\mathbf{x}_{\text{MAP}}$.

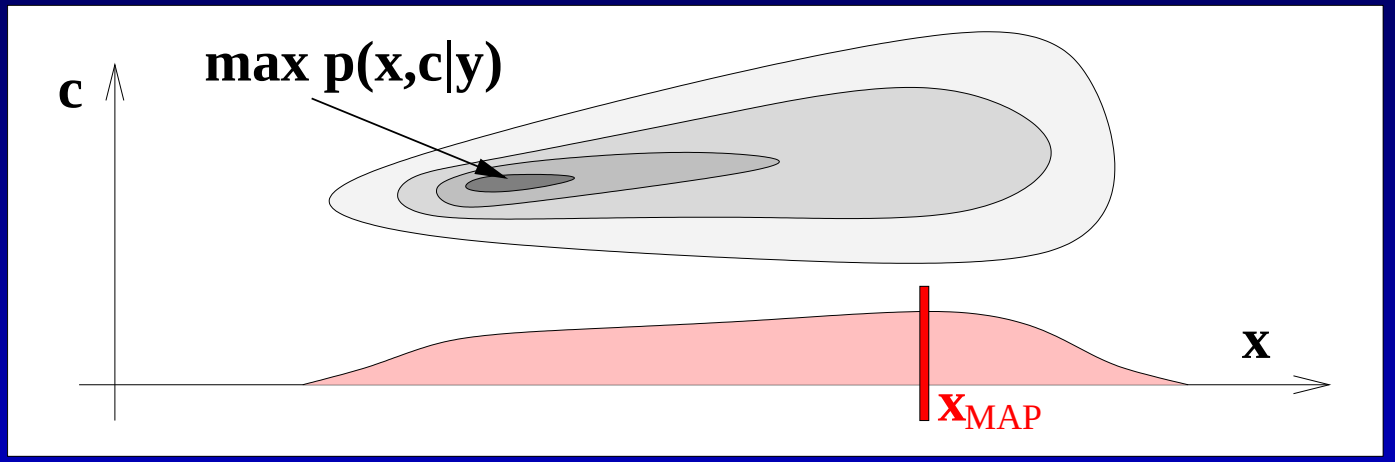
# Bayesian calibration

$\mathbf{c}$ = unknown calibration (nuisance) parameters.

Eg freckle/hair affects detector amplitude.

Want to fit them as part of inverse problem.

Standard DOT approach: find $\mathbf{c}_{\text{best fit}}$, then $\mathbf{x}_{\text{MAP}}$.



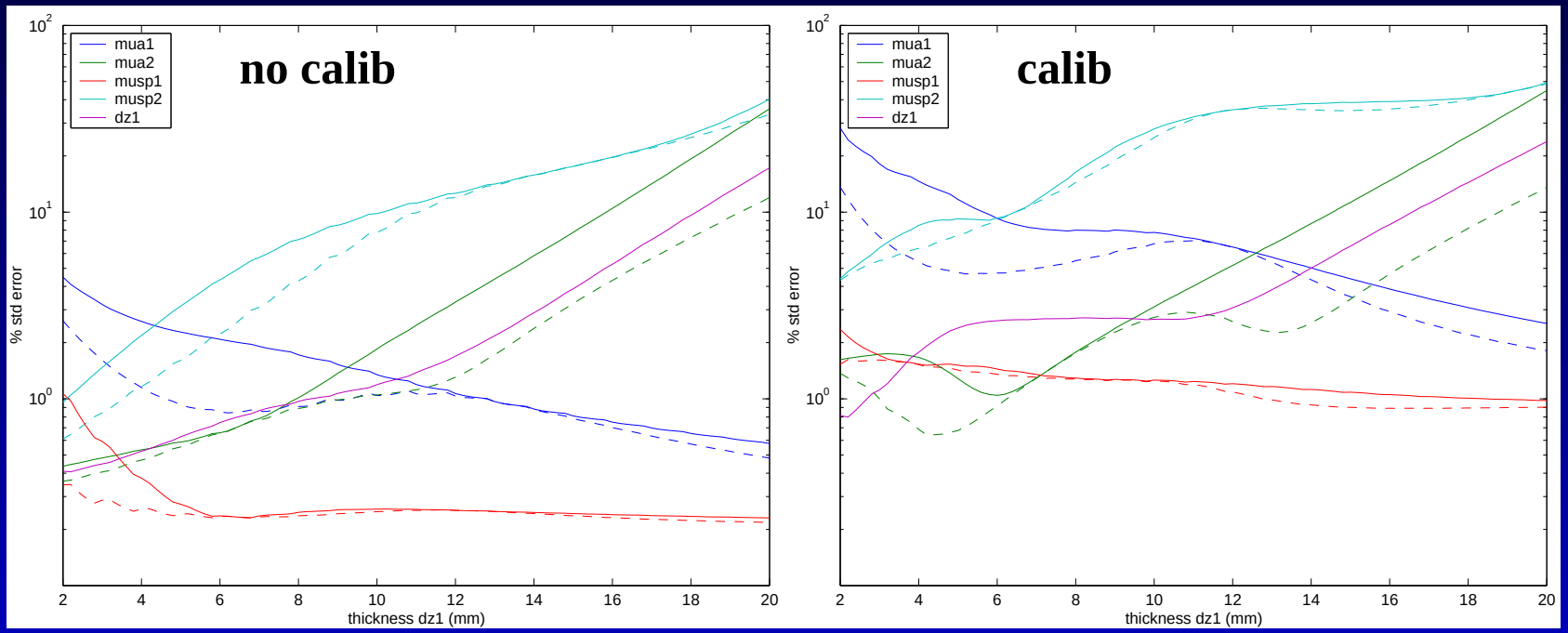

Correct: integrate $p(\mathbf{x}|\mathbf{y}) = \int p(\mathbf{x}, \mathbf{c}|\mathbf{y})\, d\mathbf{c}$.

Width in $\mathbf{c}$-space is important!

# Is MRI geometry useful?

2-layer slab, top layer thickness $\Delta z_1$.

How much does knowledge of $\Delta z_1$ improve accuracy?




1 src, 4 det, Npd=$10^5$, $\epsilon = 1\%$.

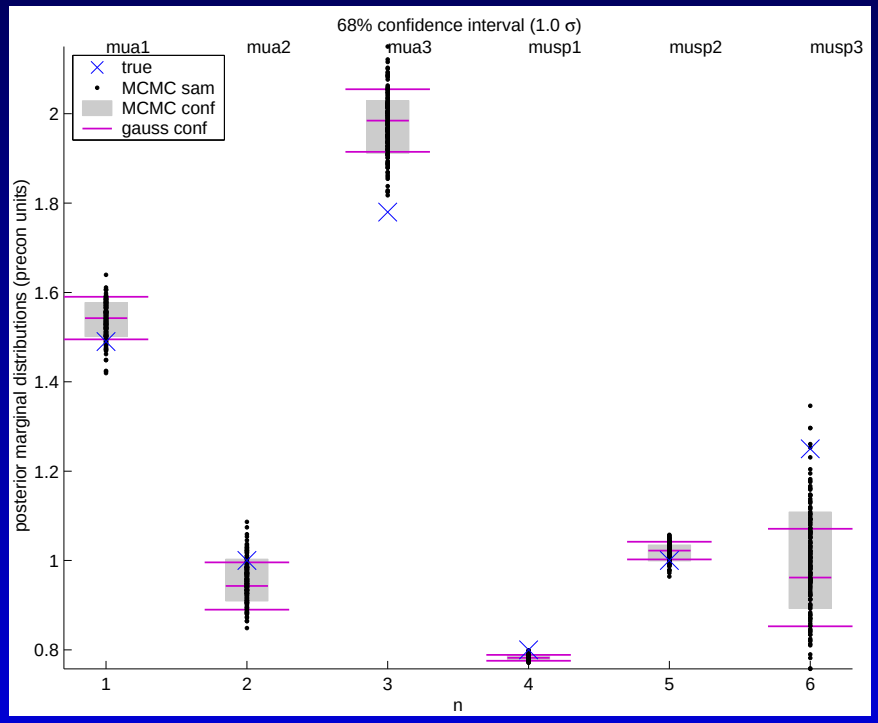
For $\Delta z_1 > 12$ mm: factor 3 improvement in $\mu_{a,2}$ error.

Std errors varying factors larger for calibrated.

# Can slab approximate the head?

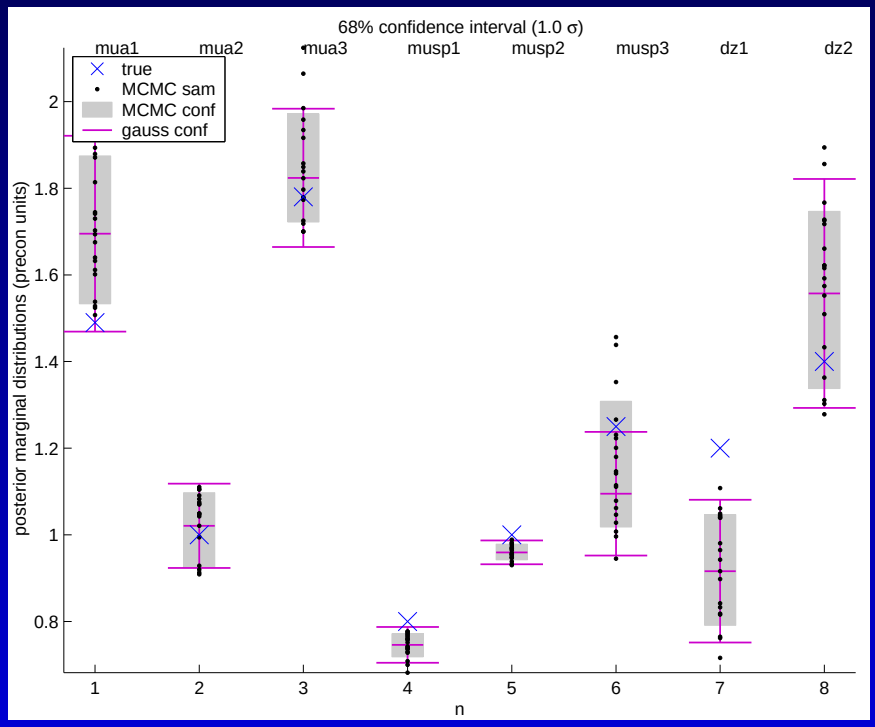
Preliminary results: can infer baseline head optical properties using 3-layer slab model ($10^2$ times faster).

$\mu_a$, $\mu_s'$ alone




brain wrong 3-$\sigma$
(since thicknesses wrong)

$\mu_a$, $\mu_s'$, & 2 thicknesses.




brain correct

# Conclusions

- Improved DOT $\rightarrow$ model complex anatomy
  - fast forward models with known errors
  - inferred baseline tissue parameters
  - slab model may be ok for head

# Conclusions

- Improved DOT $\rightarrow$ model complex anatomy
  - fast forward models with known errors
  - inferred baseline tissue parameters
  - slab model may be ok for head
- Bayes $\rightarrow$ understand full PDFs on unknowns
  - maximal use of signal information
  - predicts errorbars, correlations
  - new realistic noise model
  - statistically correct calibration

# Conclusions

- Improved DOT $\rightarrow$ model complex anatomy
  - fast forward models with known errors
  - inferred baseline tissue parameters
  - slab model may be ok for head
- Bayes $\rightarrow$ understand full PDFs on unknowns
  - maximal use of signal information
  - predicts errorbars, correlations
  - new realistic noise model
  - statistically correct calibration
- Directions:

  experimental data in progress

# Conclusions

- Improved DOT $\rightarrow$ model complex anatomy
  - fast forward models with known errors
  - inferred baseline tissue parameters
  - slab model may be ok for head
- Bayes $\rightarrow$ understand full PDFs on unknowns
  - maximal use of signal information
  - predicts errorbars, correlations
  - new realistic noise model
  - statistically correct calibration
- Directions:

  experimental data in progress

  larger $N$ optimization, adjoint method for $J$

# Conclusions

- Improved DOT $\rightarrow$ model complex anatomy
  - fast forward models with known errors
  - inferred baseline tissue parameters
  - slab model may be ok for head
- Bayes $\rightarrow$ understand full PDFs on unknowns
  - maximal use of signal information
  - predicts errorbars, correlations
  - new realistic noise model
  - statistically correct calibration
- Directions:

  experimental data in progress

  larger $N$ optimization, adjoint method for $J$

  understand effects of CSF clear layer

# Conclusions

- Improved DOT $\rightarrow$ model complex anatomy
  - fast forward models with known errors
  - inferred baseline tissue parameters
  - slab model may be ok for head
- Bayes $\rightarrow$ understand full PDFs on unknowns
  - maximal use of signal information
  - predicts errorbars, correlations
  - new realistic noise model
  - statistically correct calibration
- Directions:

  experimental data in progress

  larger $N$ optimization, adjoint method for $J$

  understand effects of CSF clear layer

  optimize S/D placement, model competition...