Advanced Applications of Synthetic MR and MAGiC

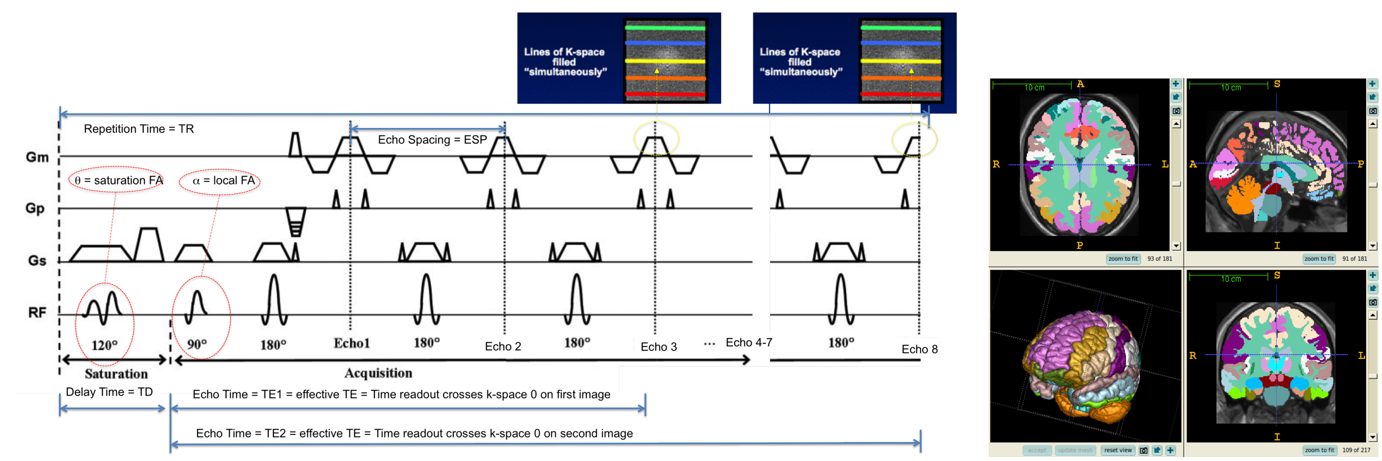
May 19, 2016

1 Pending Tasks

- @kenphwang can we get the current code for this recon.
- @kenphwang Can you provide an example data set ? is this real/imaginary data?
- @fuentesdt Code Eq. (6)
- @fuentesdt generate Figure 2

2 Problem Statement

Consider a magnetization signal M_{TD} that is defined as as function of **acquisition parameters** $\mathcal{K} = \{T_R, T_D, \theta, T_E, \alpha\}$ and **tissue properties** $\mathcal{P} \equiv \{T_1, T_2, M_0\}$. Within the scope of this project, T_R=4sec, θ=120^o, and α=90^o are **fixed**. Delay time, T_D, and echo time T_E are parameters under consideration for acquisition optimization.



(a) (b)

Figure 1: (a) Synthetic MR Pulse sequence. (b) Neuroimaging basis.

$$M_{TD}(\mathcal{K}, \mathcal{P}, \mathbf{x}) = M_0 e^{i\psi(\mathbf{x})} \left((\mathbf{x}) \frac{1 - (1 - \cos\theta) e^{-\frac{T_D}{T_1(\mathbf{x})}} - \cos\theta e^{-\frac{T_D}{T_1(\mathbf{x})}}}{1 - \cos\theta e^{-\frac{T_R}{T_1(\mathbf{x})}} cos\alpha} \right) e^{-\frac{T_E}{T_2(\mathbf{x})}} \tag{1}$$

Here, M_0 is the unsaturated magnetization, $\psi(\mathbf{x})$ is the measured phase offset, θ represents the *saturation* flip angle, and T_R and T_E denote repetition time and echo time, respectively. Parameters T_1 and T_2 represents relaxation times, and α is the *local* excitation flip angle. In general, excitation pulse α is a function of flip angle, i.e. $\alpha = \alpha(\theta)$ (@kenphwang why is this?). Note that the unsaturated magnetization M_0, along with relaxation times T_1 and T_2, are a function of spatial coordination $\mathbf{x}$. Basis functions ϕ_i represent the neuroanatomy. For completeness, consider $\phi_1 = \phi_{\text{gm}}$, $\phi_2 = \phi_{\text{wm}}$, $\phi_3 = \phi_{\text{csf}}$, $\phi_4 = \phi_{\text{tumor}}$ as a simplified set of the regions illustrated in Figure 3(b).

$$T_1(\mathbf{x}) = \sum_{i=1}^{N=4} T1_i \phi_i(\mathbf{x}) \qquad T_2(\mathbf{x}) = \sum_{i=1}^{N=4} T2_i \phi_i(\mathbf{x}) \qquad M_0(\mathbf{x}) = \sum_{i=1}^{N=4} M0_i \phi_i(\mathbf{x})$$

$$\bigcup_{i=1}^{N=4} \Omega_i = \Omega \qquad \Omega_n \cap \Omega_m = \varnothing \qquad \phi_i(\mathbf{x}) = \begin{cases} 1 & x \in \Omega_i \\ 0 & \text{otherwise} \end{cases}$$

Assume that the signal model for M_{TD} (1) is our measurment model in **image space** and is polluted with a white noise ν (with mean zero and variance $\mathbf{R}$). Hence, Eq. (1) can be written as:

$$z(\mathcal{K}, \mathcal{P}) = \underbrace{M_{TD}(\mathcal{K}, \mathcal{P}, \mathbf{x})}_{h(\mathcal{K},\mathcal{P})} + \nu \qquad \nu \in \mathbb{R}^2 \quad \mathbf{R} = \begin{bmatrix} \sigma_r & 0 \\ 0 & \sigma_i \end{bmatrix} \tag{2}$$

Note that the observation z is a function of control parameters $\mathcal{K}$ and parameters of interest $\mathcal{P}$. The ultimate goal is to provide accurate estimate of the parameters $\mathcal{P}$, given some measurements z. Precise estimation of parameters $\mathcal{P}$ crucially depends on the values of control parameters $\mathcal{K} = \{T_D, T_E\}$ ($\{T_R, \theta, \alpha\}$ **fixed**). In other words, to ensure performance of the estimation algorithm, one needs to select the control parameters $\mathcal{K} = \{T_D, T_E\}$ such that the observation z provides useful information about the parameters $\mathcal{P}$. This is achieved my maximizing the mutual information between the measurements z and parameters of interest $\mathcal{P}$. Within this framework we will consider the tissue properties to be normally distributed Gaussian parameters.

$$\mathrm{T1}_{\mathrm{WM}} = \mathcal{N}(100ms, 20ms) \qquad \mathrm{T1}_{\mathrm{GM}} = \mathcal{N}(120ms, 20ms) \qquad \mathrm{T1}_{\mathrm{CSF}} = \mathcal{N}(320ms, 20ms) \qquad \mathrm{T1}_{\mathrm{Tumor}} = \mathcal{N}(300ms, 20ms)$$

$$\mathrm{T2}_{\mathrm{WM}} = \mathcal{N}(100ms, 20ms) \qquad \mathrm{T2}_{\mathrm{GM}} = \mathcal{N}(120ms, 20ms) \qquad \mathrm{T2}_{\mathrm{CSF}} = \mathcal{N}(320ms, 20ms) \qquad \mathrm{T2}_{\mathrm{Tumor}} = \mathcal{N}(300ms, 20ms)$$

$$\mathrm{M0}_{\mathrm{WM}} = \mathcal{N}(100??, 20??) \qquad \mathrm{M0}_{\mathrm{GM}} = \mathcal{N}(120??, 20??) \qquad \mathrm{M0}_{\mathrm{CSF}} = \mathcal{N}(320??, 20??) \qquad \mathrm{M0}_{\mathrm{Tumor}} = \mathcal{N}(300??, 20??)$$

(@kenphwang need exact numbers)

3 Optimal (T_D,T_E) Design

As discussed before, performance of estimation process crucially depends on the value of control parameters $\mathcal{K}$. Hence, it is important to develop mathematical tools to identify the control parameters $\mathcal{K}$ such that they provide the best observation data for accurate estimation of parameter $\mathcal{P}$. This is equivalent with maximizing the mutual information between the observation data and parameters $\mathcal{P}$. Based on information theory, mutual information is defined as the reduction of uncertainty in one parameter due to knowledge of the other parameter.

$$I(\mathcal{P}; z) = \int_z \int_p p(\mathcal{P}, z) \ln \left(\frac{p(\mathcal{P}, z)}{p(\mathcal{P})p(z)} \right) d\mathcal{P} dz \tag{3}$$

We make use of Bayes theorem to simplify the above equation. By substituting $p(\mathcal{P}, z)$ with $p(z|\mathcal{P})p(\mathcal{P})$, Eq. (3) can be written as:

$$I(\mathcal{P}; z) = \int_z \int_p p(z|\mathcal{P})p(\mathcal{P}) \ln \left(\frac{p(z|\mathcal{P})p(\mathcal{P})}{p(\mathcal{P})p(z)} \right) d\mathcal{P} dz \tag{4}$$

Or,

$$I(\mathcal{P}; z) = \int_z \int_p p(z|\mathcal{P})p(\mathcal{P}) \ln \left[p(z|\mathcal{P}) \right] d\mathcal{P} dz - \int_z p(z) \ln p(z) dz \tag{5}$$

Note that due to dependence of observation data z on control parameters, the mutual information $I(\mathcal{P}; z)$ is a function of control parameter $\mathcal{K}$. In order to maximize the reduction of uncertainty in parameter estimate (i.e. to have the most confident estimates of the parameter $\mathcal{P}$), one can simply maximize the mutual information between the observation data and parameters of interest:

$$\max_{\mathcal{K} \in \mathcal{F}} I(\mathcal{P}; z) \qquad \mathcal{K} = \{T_D, T_E\} \in \mathcal{F} = (0, 4sec] \times (0, 140ms] \subset \mathbb{R}^2 \tag{6}$$

The above maximization results in *optimal* values of control parameter $\mathcal{K}$ for accurate estimation of parameter $\mathcal{P}$. The feasible set, $\mathcal{F}$, is determined by the pulse sequence acquisition physics. Note that in Eq. (5), $p(z|\mathcal{P})$ is defined as a Gaussian distribution with mean $h(\mathcal{K}, \mathcal{P})$ and variance $\mathbf{R}$. As well, $p(\mathcal{P})$ denotes the prior distribution of parameter $\mathcal{P}$, which for the ease of calculations, is considered to be a Gaussian distribution with some prior mean $\hat{\mathcal{P}}^-$ and prior covariance Σ^-, i.e. $p(\mathcal{P}) \sim \mathcal{N}(\hat{p}^-, \Sigma^-)$. Method of quadrature points can be used to evaluate Eq. (5).

We emphasize here that the mutual information will be the same on different pixels with the same tissue types. This is due to the similarities in statistics of $\mathcal{P}$ between the two different pixels with the same tissue properties. In other words, whenever two different pixels have the same tissue properties, then the distribution of parameter $\mathcal{P}$, denoted by $p(\mathcal{P})$, is the same and so is the value of mutual information. Hence, there is no need to evaluate the mutual information for each pixel in a region with the same tissue type. On the other hand, in a case that the tissue properties for each pixel are different from the other, then the mutual information needs to be evaluated for each and every pixel of interest.

The final plot over delay time and echo time is shown in Figure 2. The current acquistion utilizes 8 total combinations of (effective) echo times and delay times, 2 echo times and 4 delay times.

The information content of this acquisition scheme is calculated as the superposition

$$I^{\text{current}} = \sum_{i=1}^{8} I((T_D, T_E)_i) \qquad \text{(@kenphwang need exact numbers)}$$

$(T_D, T_E) = \{$(20ms, 15ms),(40ms, 15ms),(70ms, 15ms),(90ms, 15ms), (20ms, 35ms),(40ms, 35ms),(70ms, 35ms),(90ms, 35ms)$\}$

An optimal acquisition strategy either maintains or improves this information content with less samples to (1) improve time and (2) accuracy.

$$I^{\text{current}} < I^{\text{optimal}} = \sum_{i=1}^{M \leq 8} I((T_D^*, T_E^*)_i)$$

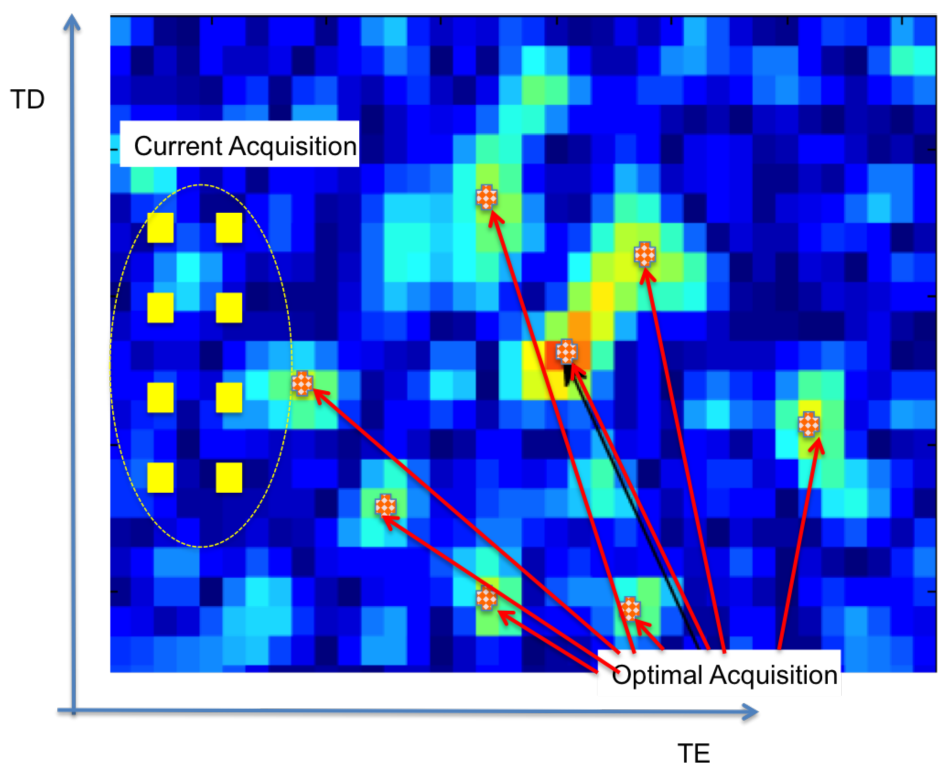


Figure 2: Information content as function of $(T_D, T_E) \in \mathcal{F} = (0, 4sec] \times (0, 140ms] \subset \mathbb{R}^2$. The feasible set, $\mathcal{F}$, is determined by the pulse sequence acquisition physics.

3.1 MI Approximations: Maximizing the Variance

Unfortunately, evaluation of mutual information and solving the above optimization problem is computationally intractable for most of the practical applications. Hence, one needs to simplify the original optimization problem and find the approximate solution for optimal location of each observation. A simpler alternative in finding useful

$k-$space locations is to maximize the entropy (uncertainty) in model outputs $\mathcal{U}$, instead of maximizing the mutual information $I(\mu; z)$ [?]. Based on information theory, the entropy of $\mathcal{U}$, denoted by $h(\mathcal{U})$ is defined as:

$$h(\mathcal{U}) = -\int_{\mathcal{U}} ln\left(p(\mathcal{U})\right) p(\mathcal{U}) d\mathcal{U} \tag{7}$$

where, $p(\mathcal{U})$ is the probability density function of $\mathcal{U}$. One can simplify the problem by just maximizing the entropy of model outputs, instead of maximizing the mutual information. We emphasize here that this simplification may have its drawback in not giving the *globally* optimal locations on $k-$space, but as we will show in the following, it will result in great simplification of the problem.

To proceed with maximizing the entropy $h(\mathcal{U})$, we first note that based on Maximum Entropy Principle [?], probability density function $p(\mathcal{U})$ can be parametrized in terms of its central moments as

$$p(\mathcal{U}) = \lim_{N\to\infty} \left(e^{\sum_{n=0}^{N} \lambda_n (\mathcal{U}-\hat{\mathcal{U}})^n} \right), \quad \lambda_n \in \mathbb{R} \tag{8}$$

where, $\hat{\mathcal{U}}$ is provided by (**??**). By substituting Eq. (8) in Eq. (7), we will have:

$$\begin{aligned} h(\mathcal{U}) = -\int_{\mathcal{U}} \lim_{N\to\infty} \left(\sum_{n=0}^{\infty} \lambda_n (\mathcal{U}-\hat{\mathcal{U}})^n \right) p(\mathcal{U}) d(\mathcal{U}) \\ = \lim_{N\to\infty} (-\lambda_0 - \sum_{n=2}^{N} \lambda_n m_n) \end{aligned} \tag{9}$$

where, m_ns are the central moments of $\mathcal{U}$, defined by Eq. (**??**). Hence, entropy of $p(\mathcal{U})$ can be described in terms of its central moments, as shown in Eq. (9). Now, by approximating $p(\mathcal{U})$ with its first two moments and truncating the above expansion (i.e. letting $N = 2$), we have

$$h(\mathcal{U}) \simeq -\lambda_0 - \lambda_2 m_2, \quad \lambda_0, \lambda_2 \in \mathbb{R} \tag{10}$$

Therefore, to maximize the entropy one can only maximize the variance, i.e.

$$\max_K h(\mathcal{U}) = -\lambda_0 - \lambda_2 \max_K (m_2), \quad \lambda_0, \lambda_2 \in \mathbb{R} \tag{11}$$

subject to:

$$|K^i - K^j| \geq N_d, \quad \forall\, i, j \in 1, 2, \cdots, N, \quad i \neq j \tag{12}$$

where, $|K^i - K^j| = |(k_x^i, k_y^i, k_z^i) - (k_x^j, k_y^j, k_z^j)|$ denotes the distance between the i^{th} and j^{th} locations on $k-$space and $m_2 = Var(\mathcal{U})$ is defined from Eq. (**??**). Hence, the locations on $k-$space with the highest value of variance for $\mathcal{U}$ are a good approximation of optimal locations for data observations.

Eq. (12) is considered to ensure that every distinct pair of $k-$space observations are at least by a distance N_d apart from each other. This constraint is used in order to compensate possible dependencies that can be introduced due to approximation of the original problem. We assumed $N_d = 1$ through this manuscript.

One should note that different order of truncation (i.e. different values of N) can be used to approximate the entropy in Eq. (9). Clearly, higher order approximation (greater values of N) leads to more accurate approximation of entropy. However, the downside of using higher order terms is that one needs to find the corresponding λ_n coefficients for each term. Finding corresponding λ_n coefficients requires solving an optimization problem which could increase computational cost of the whole procedure. Hence, lower order approximation of entropy is of more interest due to real time applications of the proposed technique.

The intuition behind the idea of maximizing the variance is that the points on $k-$space with lower value of variance are less sensitive to model perturbations (resulted from parameter uncertainties) and vice versa. Hence, it is better to select the points on $k-$space with highest sensitivity with respect to model uncertainties. For instance, a point with zero variance on $k-$space will not be a good candidate for data observation since no matter what the values of uncertain parameters are, model output will always be the same at that specific point. On the other hand, a point on $k-$space with large value of variance means that the model output at that location is *highly sensitive* to model uncertainties. Hence, a measurement at that $k-$space location would be of more interest. Therefore, the points with highest values of variance of $\mathcal{U}$ are *better candidates* for data observations.

(@dmitchell412 need Variance calculation algorithm compiled here with Quadrature)

3.2 MI Calculations

(@dmitchell412 need Entropy calculation algorithm compiled here using linearization about quadrature points)

A more direct calculation of the entropy is given as

$$MI = H(z) \approx$$

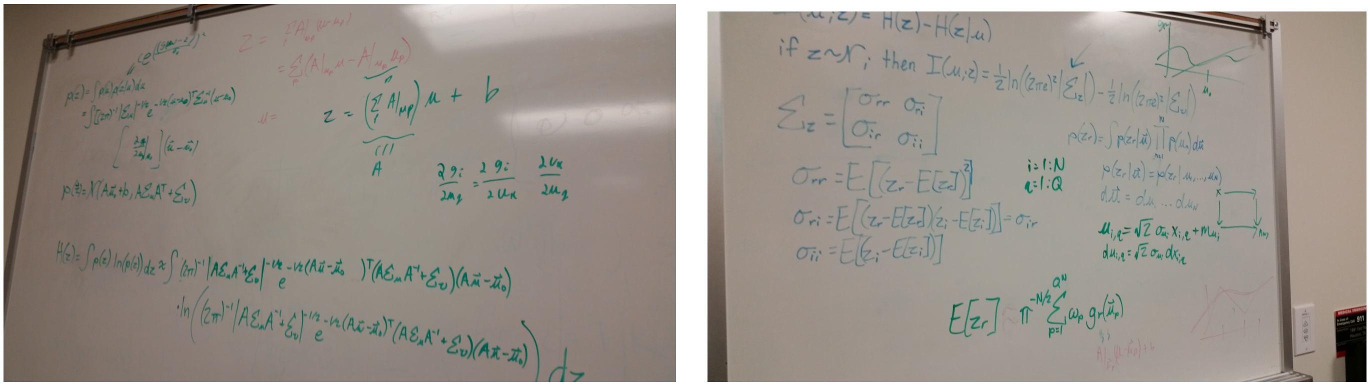


Figure 3: WIP - @dmitchell412 Need higher order entropy approximation

4 Overall Picture

The following diagram illustrates the general work-flow of the process:

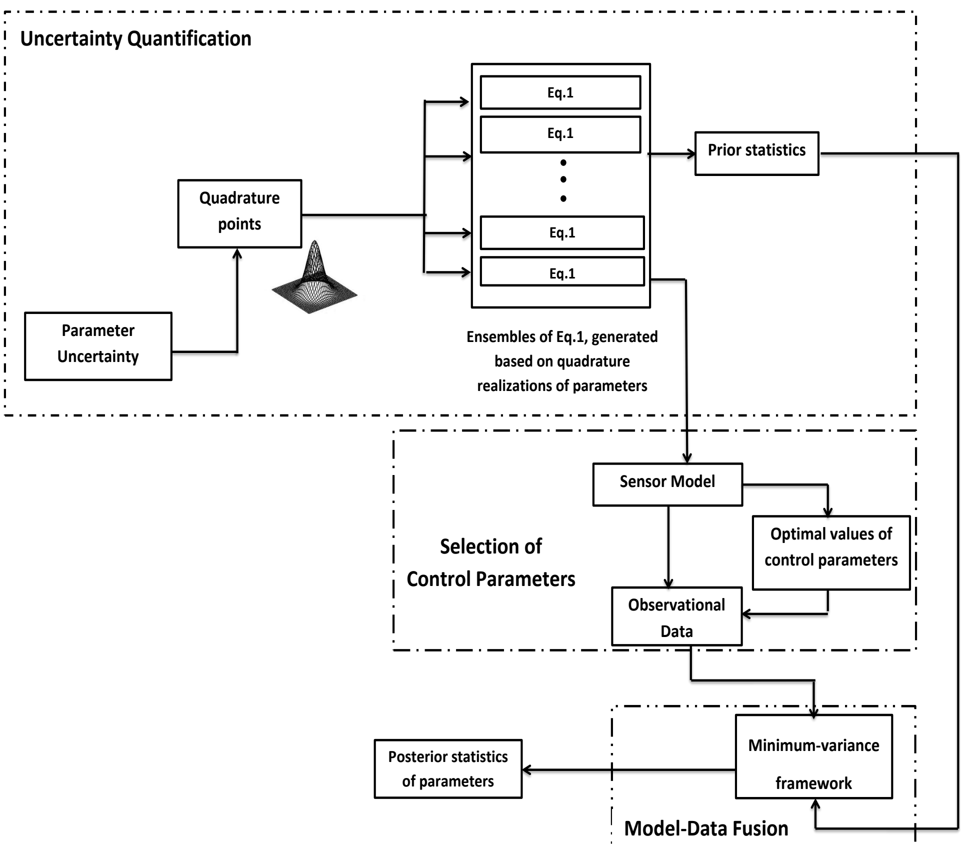


Figure 4: Schematic view of the estimation process

5 WIP - Inverse Problem Framework

$$\vec{z} = \mathcal{G}(\theta) + \eta \qquad \eta \sim \mathcal{N}(0, \Sigma_z)$$
$$p(z|\theta) = \exp\left(\|\vec{z} - \mathcal{G}(\theta)\|^2_{\Sigma_z}\right)$$
$$d\left(\vec{z}, \mathcal{G}(\theta^*)\right) = \min_{\theta \in \Omega} d\left(\vec{z}, \mathcal{G}(\theta)\right) \qquad \theta = (\mu_{\text{CSF}}, \mu_{\text{GM}}, \mu_{\text{WM}}, \mu_{\text{Tumor}})$$

6 WIP - T1, T2, M0 Reconstruction

Given the image space date for multiple acquistion parameters $\{M_{TD}(\mathcal{K}_1), M_{TD}(\mathcal{K}_2), M_{TD}(\mathcal{K}_2), ...\}$, $\mathcal{K}_i = \{T_{R_i}, T_{D_i}, \theta_i, T_{E_i}, \alpha_i\}$ (@kenphwang 2 delay times and 4 echoes correct?),

The reconstruction algorithms for T1, T2, M0 is as follows:

-
-
-

7 WIP - k-space signal model

@wstefan @kenphwang need to develop full k-space model of echo train trajectories for subsampling.

Consider the schematic of the pulse sequence for the acquistion provided in Figure ??(a). For a given tissue, $\Omega \subset \mathbb{R}^3$, with heterogenious intrinistic relaxation properties $T1(r)$ and $T2^*(r)$ we will assume a multi-echo signal, $s_l(t,n) \in \mathbb{C}$, from the l-th coil at the n-th echo of the form of a linear combination of damped exponentials. The signal strength is dependent on the coil sensitivity $c_l(r)$.

$$\begin{aligned} s_l(t,n) &= \int_\Omega c_l(r) w[n] e^{-i\int_0^t \gamma \vec{G}(\tau)\cdot r d\tau}\, dr \qquad w[n] = \sum_j^{N_{\text{species}}} C_j \Lambda_j^n \qquad n = 0, 1, 2, \ldots, N_{\text{echo}} - 1 \\ \Lambda_j^n &= e^{-\left(\underbrace{i\ \Delta B_{0_j}(r)}_{\text{off-resonance}} + \frac{1}{T2_j^*}\right)(\text{TE}+n\ \text{ESP})} \qquad C_j = \frac{M_{0_j}\sin(\gamma\theta_N)\left(1 - e^{-TR/T1_j}\right)}{\left(1 - \cos(\gamma\theta_N)\, e^{-TR/T1_j}\right)} e^{-i\phi_j} \in \mathbb{C} \end{aligned} \tag{13}$$

Each weighted exponential, $C_j\Lambda_j$, represents a distinct chemical species, ie water, fat, sodium hydroxide, etc. The complex amplitude, $C_j \in \mathbb{C}$, depends on (1) acquisition parameters - repetition time, TR, and flip angle, θ_N, as well as (2) tissue dependent properties - spin lattice relaxation, $T1_j$, proton density M_{0_j}, and initial phase offset, ϕ_j. Similarly, the complex exponential, Λ_j, depends on (1) acquisition parameters - echo time, TE, echo spacing , ESP, as well as (2) tissue dependent properties - spin spin relaxation, $T2_j^*$, and tissue depending off-resonsance arising from temperature change, motion, and/or susceptibility, $\Delta B_{0_j}(r)$.

$$\Delta B_{0_j}(r) = \begin{cases} 2\pi f_j = 2\pi \underbrace{\alpha B_0 \Delta T_j(r)}_{f_j} & \text{temperature} \\ \quad\quad ... = & \text{susceptibility} \\ \quad\quad ... = & \text{motion} \end{cases}$$

We will assume:

- off-resonance, ΔB_{0_j} does not depend on readout time
- constant gradients

$$\begin{cases} k_x = \gamma G(t - TE) \\ k_y = \gamma G_y T_{pe} \\ k_z = \gamma G_z T_{pe} \end{cases} \qquad |t - TE| < T_{acq}/2 \qquad \vec{k} \equiv \frac{\gamma\, \vec{G}\, t}{2\pi}$$

The time in this signal model, t, represents the measurements during the readout and is less the the repetition time $t < TR \approx 500ms$.

$$t = iii \cdot \Delta t \quad iii = 0, ..., 255 \qquad \Delta t = \frac{T_{acq}}{256}$$

Note that the acquisition time for a single echo/single slice under this model may be estimated as the number of phase encodes times the repetition time.

$$\text{acquisition time} = \#\text{phase encodes} \ \cdot TR \leq 256 \cdot 500ms \approx 2\text{min}$$

- the phase induced temperature change is measured at the echo time

$$\int_0^{\text{TE}} f(\tau) d\tau \approx f(TE)TE$$

Under these assumptions, the measurements at the n-th echo $s(t_i, n)$, have the intuitive interpretation as the Fourier coefficient of the complex image, $c_l(r) \sum_j C_j \Lambda_j^n : \mathbb{R}^3 \to \mathbb{C}$.

$$s_l(t, n) = \text{Fourier transform} \left\{ \underbrace{c_l(r) \sum_j C_j \Lambda_j^n}_{\equiv f(r):\mathbb{R}^3 \to \mathbb{C}} \right\} = \mathcal{F}(f(r)) = \int_\Omega \left(c_l(r) \sum_j C_j \Lambda_j^n \right) e^{-2\pi i \vec{k} \cdot r} \, dr = s_l(k_x, k_y, k_z, n) \tag{14}$$

TODO - @dmitchell412 - derive this model and verify

Within the context/notation of Eqn (23),

$$s_l(k_x, k_y, k_z, n) = \mathcal{G}\left(\underbrace{\vec{k}, TE, TR, ESP, N_{echo}, N_{species}, \theta_N}_{k}, \underbrace{T1, T2, \Delta B_0}_{\theta} \right)$$

References

[Ahmed and Gokhale, 1989] Ahmed, N. A. and Gokhale, D. (1989). Entropy expressions and their estimators for multivariate distributions. *IEEE Transactions on Information Theory*, 35(3):688–692.

[Carp et al., 2004] Carp, S. A., Prahl, S. a., and Venugopalan, V. (2004). Radiative transport in the delta-P1 approximation: accuracy of fluence rate and optical penetration depth predictions in turbid semi-infinite media. *Journal of biomedical optics*, 9(3):632–47.

[Duck, 1990] Duck, F. (1990). Physical properties of tissue: a comprehensive reference book. *London, UK: Academic.*

[Durrett, 2010] Durrett, R. (2010). *Probability: theory and examples.* Cambridge university press.

[Fahrenholtz et al., 2015] Fahrenholtz, S., Moon, T., Franco, M., Medina, D., Hazle, J. D., Stafford, R. J., Maier, F., Danish, S., Gowda, A., Shetty, A., Warburton, T., and Fuentes, D. (2015). A Model Evaluation Study for Treatment Planning of Laser Induced Thermal Therapy. *International Journal of Hyperthermia.* in preparation.

[Fahrenholtz et al., 2013] Fahrenholtz, S., Stafford, R. J., Hazle, J., and Fuentes, D. (2013). Generalised polynomial chaos-based uncertainty quantification for planning MRgLITT procedures. *International Journal of Hyperthermia*, 29(4):324–335. PMC3924420.

[Fasano et al., 2010] Fasano, A., Hömberg, D., and Naumov, D. (2010). On a mathematical model for laser-induced thermotherapy. *Applied Mathematical Modelling*, 34(12):3831–3840.

[Fuentes et al., 2013] Fuentes, D., Elliott, A., Weinberg, J. S., Shetty, A., Hazle, J. D., and Stafford, R. J. (2013). An Inverse Problem Approach to Recovery of In-Vivo Nanoparticle Concentrations from Thermal Image Monitoring of MR-Guided Laser Induced Thermal Therapy. *Ann. BME.*, 41(1):100–111. PMC3524364.

[Fuentes et al., 2010] Fuentes, D., Feng, Y., Elliott, A., Shetty, A., McNichols, R. J., Oden, J. T., and Stafford, R. J. (2010). Adaptive Real-Time Bioheat Transfer Models for Computer Driven MR-guided Laser Induced Thermal Therapy. *IEEE Trans. Biomed. Eng.*, 57(5). Cover Page, PMC3857613.

[Fuentes et al., 2009] Fuentes, D., Oden, J. T., Diller, K. R., Hazle, J., Elliott, A., , Shetty, A., and Stafford, R. J. (2009). Computational Modeling and Real-Time Control of Patient-Specific Laser Treatment Cancer. *Ann. BME.*, 37(4):763. PMC4064943.

[Fuentes et al., 2011] Fuentes, D., Walker, C., Elliott, A., Shetty, A., Hazle, J., and Stafford, R. J. (2011). MR Temperature Imaging Validation of a Bioheat Transfer Model for LITT. *International Journal of Hyperthermia*, 27(5):453–464. Cover Page, PMC3930085.

[MacLellan et al., 2013] MacLellan, C. J., Fuentes, D., Elliott, A. M., Schwartz, J., Hazle, J. D., and Stafford, R. J. (2013). Estimating nanoparticle optical absorption with magnetic resonance temperature imaging and bioheat transfer simulation. *International Journal of Hyperthermia*, (0):1–9.

[Madankan et al., 2015] Madankan, R., Stefan, W., Hazle, J. D., Stafford, R. J., and Fuentes, D. (2015). Accelerated Model-based Signal Reconstruction for Magnetic Resonance Imaging in Presence of Uncertainties. *IEEE Trans. Med. Img.* submitted.

[Martin et al., 2012] Martin, J., Wilcox, L. C., Burstedde, C., and Ghattas, O. (2012). A stochastic newton mcmc method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487.

[Modest, 2013] Modest, M. F. (2013). *Radiative heat transfer*. Academic press.

[Needles et al., 2010] Needles, A., Heinmiller, A., Ephrat, P., Bilan-Tracey, C., Trujillo, A., Theodoropoulos, C., Hirson, D., and Foster, F. (2010). Development of a combined photoacoustic micro-ultrasound system for estimating blood oxygenation. In *Ultrasonics Symposium (IUS), 2010 IEEE*, pages 390–393. IEEE.

[Stefan et al., 2015] Stefan, W., Fuentes, D., Yeniaras, E., Hwang, K., Hazle, J. D., and Stafford, R. J. (2015). Novel Method for Background Phase Removal on MRI Proton Resonance Frequency Measurements. *Trans. Medical Imaging*. in review.

[Welch, 1984] Welch, A. J. (1984). The thermal response of laser irradiated tissue. *Quantum Electronics, IEEE Journal of*, 20(12):1471–1481.

[Welch and Van Gemert, 1995] Welch, A. J. and Van Gemert, M. J. (1995). *Optical-thermal response of laser-irradiated tissue*, volume 1. Springer.

[Wray et al., 1988] Wray, S., Cope, M., Delpy, D. T., Wyatt, J. S., and Reynolds, E. O. R. (1988). Characterization of the near infrared absorption spectra of cytochrome aa3 and haemoglobin for the non-invasive monitoring of cerebral oxygenation. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 933(1):184–192.

[Yung et al., 2015] Yung, J., Fuentes, D., MacLellan, C. J., Maier, F., Hazle, J. D., and Stafford, R. J. (2015). Referenceless Magnetic Resonance Temperature Imaging using Gaussian Process Modeling. *Medical Physics*. in review.

A Bayes - An intuitive example

Bayes theorem is fundamental to the approach and is immediately follows from the definition of conditional probability

$$\left.\begin{aligned} p(y|x) &\equiv \frac{p(x,y)}{p(x)} \\ p(x|y) &\equiv \frac{p(x,y)}{p(y)} \end{aligned}\right\} \Rightarrow p(y|x)p(x) = p(x,y) = p(x|y)p(y) \Rightarrow \qquad \begin{aligned} p(y|x) &= \frac{p(x|y)p(y)}{p(x)} \\ p(x|y) &= \frac{p(y|x)p(x)}{p(y)} \end{aligned}$$

As a concrete example, consider the explicit two dimensional joint Gaussian distribution as a medium for understanding. Here we have two random variables $\mathbf{x}_1$ and $\mathbf{x}_2$ defined on the same probability space, Ω.

$$\mathbf{x}_i : \Omega \to \mathbb{R} \qquad P\left(\{\omega : \mathbf{x}_i(\omega) \in A\}\right) = \int_A p(\eta_i) d\eta_i$$

Intuitively, if we are **given** the joint distribution, $p(\eta_1, \eta_2)$, knowledge of the realization of one particular random variable provides information on the realization of the second random variable.

$$p(\eta_1, \eta_2) = \frac{1}{2\ \pi\ \sqrt{\det \Sigma}} \exp\left(\frac{1}{2} \begin{bmatrix} \eta_1 - \mu_1 \\ \eta_2 - \mu_2 \end{bmatrix}^\top \underbrace{\begin{bmatrix} \sigma_1^2 & r_{12}\sigma_1\sigma_2 \\ r_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}}_{\equiv \Sigma} \begin{bmatrix} \eta_1 - \mu_1 \\ \eta_2 - \mu_2 \end{bmatrix} \right)$$

See [?] (Sec 3.10), characteristic functions are used to show that individual marginal densities of joint Gaussian random variable is also Gaussian.

$$p(\eta_1) = \int_{\eta_2} p(\eta_1, \eta_2)\ d\eta_2 = \frac{1}{\sqrt{2\ \pi\ \sigma_2^2}} \exp\left(-\frac{(\eta_1 - \mu_1)^2}{2\sigma_1^2} \right)$$

$$p(\eta_2) = \int_{\eta_1} p(\eta_2, \eta_1)\ d\eta_1 = \frac{1}{\sqrt{2\ \pi\ \sigma_1^2}} \exp\left(-\frac{(\eta_2 - \mu_2)^2}{2\sigma_2^2} \right)$$

Conditional probablity is *defined* through the algegraic reduction of the ratio of the joint and the marginal densities

$$p(\eta_1|\eta_2) = \frac{p(\eta_1, \eta_2)}{p(\eta_2)} = \frac{1}{\sqrt{2\ \pi\ \sigma_{1|2}^2}} \exp\left(-\frac{(\eta_1 - \mu_{1|2})^2}{2\sigma_{1|2}^2} \right) = \frac{p(\eta_1)}{p(\eta_2)} \frac{1}{\sqrt{2\ \pi\ \sigma_{2|1}^2}} \exp\left(-\frac{(\eta_2 - \mu_{2|1})^2}{2\sigma_{2|1}^2} \right)$$

$$\mu_{1|2} = \mu_1 - \frac{r_{12}\sigma_1\sigma_2}{\sigma_2^2}(\eta_2 - \mu_2) \qquad \sigma_{1|2} = \sigma_1^2 - \frac{(r_{12}\sigma_1\sigma_2)^2}{\sigma_2^2}$$

$$\mu_{2|1} = \mu_2 - \frac{r_{12}\sigma_1\sigma_2}{\sigma_1^2}(\eta_1 - \mu_1) \qquad \sigma_{2|1} = \sigma_2^2 - \frac{(r_{12}\sigma_1\sigma_2)^2}{\sigma_1^2}$$

B WIP - Model Data Fusion

After finding the optimal values of the control parameter $\mathcal{K}$, we can proceed and perform the model-data fusion to get a better understanding about the uncertainties involved in parameters $\mathcal{P}$. The fusion of observational data with mathematical model predictions promises to provide greater understanding of physical phenomenon than either approach alone can achieve. In here, a minimum variance framework is being used for model - data fusion. Based on minimum variance technique, posterior statistics of parameter $\mathcal{P}$ can be written as:

$$\hat{\mathcal{P}}^+ = \hat{\mathcal{P}}^- + \mathbf{K}[z - \underbrace{\mathcal{E}^-[h(\mathcal{K}, \mathcal{P})]}_{h^-}] \tag{15}$$

$$\Sigma^+ = \Sigma^- + \mathbf{K}\Sigma_{hh}\mathbf{K}^T \tag{16}$$

where, the gain matrix K is given by

$$\mathbf{K} = \Sigma_{\mathcal{P}z} \left(\Sigma_{hh}^- + \mathbf{R} \right)^{-1} \tag{17}$$

Here, $\hat{\mathcal{P}}^-$ and $\hat{p}^+$ represent prior and posterior values of the mean for parameter vector $\mathcal{P}$, respectively:

$$\hat{\mathcal{P}}^- \equiv \mathcal{E}^-[\mathcal{P}] = \int \mathcal{P}^- p(\mathcal{P}) d\mathcal{P} \hat{\mathcal{P}}^+ \equiv \mathcal{E}^+[\mathcal{P}] = \int \mathcal{P}^+ p(\mathcal{P}) d\mathcal{P} \tag{18}$$

where, $p(\mathcal{P})$ denotes the probability density function of parameter $\mathcal{P}$. Similarly, the prior and posterior covariance matrices Σ^- and σ^+ can be written as:

$$\Sigma^- \equiv \mathcal{E}[(\mathcal{P} - \hat{\mathcal{P}}^-)(\mathcal{P} - \hat{\mathcal{P}}^-)^T] \tag{19}$$

$$\Sigma^+ \equiv \mathcal{E}[(\mathcal{P} - \hat{\mathcal{P}}^+)(\mathcal{P} - \hat{\mathcal{P}}^+)^T] \tag{20}$$

The matrices $\Sigma_{\mathcal{P}z}$ amd Σ_{hh} are defined as:

$$\Sigma_{\mathcal{P}z} \equiv \mathcal{E}[(\mathcal{P} - \hat{\mathcal{P}})(h - \hat{h}^-)^T] \tag{21}$$

$$\Sigma_{hh} \equiv \mathcal{E}[(h - \hat{h}^-)(h - \hat{h}^-)^T] \tag{22}$$

Eq. (15) along with Eq. (16) provide posterior mean and covariance of parameter $\mathcal{P}$ given observation data $\tilde{z}$ and model predictions $h(\mathcal{K}, \mathcal{P})$. We emphasize here that the optimal values of $\mathcal{K}$, obtained from Eq. (6), are used in Eq. (15).

C WIP - Mathematical Framework

The underlying philosophy and assumptions within our approach is that the physics models are 1st order accurate or within 70-80% of the needed accuracy and the error is adequate within the assumed Gaussian noise. Gaussian distributions provide analytical representations of the random variables of interest (ie T1, T2) within the Bayesian setting and provide a crux for understanding. In particular, we say that a random variable η belongs to a multi-variate normal distribution of mean $\mu \in \mathbb{R}^n$ and covariance $\Sigma \in \mathbb{R}^{n \times n}$

$$\eta \sim \mathcal{N}(\mu, \Sigma) \Rightarrow p(\eta) = \frac{1}{2\ \pi\ \det \Sigma} \exp\left(-\frac{1}{2}\|\mu - \eta\|^2_\Sigma\right)$$

1. Our data acquistion model, $\mathcal{G}(\vec{k}, \theta) : \mathbb{R}^a \times \mathbb{R}^m \to \mathbb{R}^n$, maps deterministic acquisition parameters, $\vec{k} \in \mathbb{R}^a$, and uncertain parameters, $\theta \in \mathbb{R}^m$ to observables, $\vec{z} \in \mathbb{R}^n$ (or $\vec{z} \in \mathbb{C}^n$). Explicitly, we will assume that the measurement models are corrupted by zero mean white noise noise of a **known** covariance matrix, $\Sigma_z \in \mathbb{R}^{n \times n}$

$$\begin{aligned} \vec{z} &= \mathcal{G}(\vec{k}; \theta) + \eta \qquad \eta \sim \mathcal{N}(0, \Sigma_z) \\ \vec{k} &= (\text{TE, TR, etc}) \\ \theta &= (\text{T1, T2, etc}) \end{aligned} \tag{23}$$

   η may be interpreted as the measurement noise or the acquisition noise in the sensor model. For a deterministic measurement model $\mathcal{G}$, the conditional probablity distribution has an explicit analytical form and may be written as a **known** Gaussian distribution.

$$p(\vec{z}|\theta) = \mathcal{N}(\mathcal{G}(\vec{k}; \theta), \Sigma_z)$$

2. Additional **known** information is the prior probability distributions for the model parameters, $p(\theta)$. For simplicity, assume that Prior parameters are Gaussian distributed of **known** mean, $\hat{\theta}$ and covariance, Σ_θ

$$\theta \sim \mathcal{N}(\hat{\theta}, \Sigma_\theta)$$

3. Bayes theorem is fundamental to the approach. The probability of the measurements $p(z)$ must be interprited in terms of the known information. The probability of the measurements may be derived from the marginalization of the joint probability and has the interpretation as the projection of the joint probability onto the measurement axis.

$$p(z) = \int_\theta p(\theta, z)\ d\theta = \int_\theta p(z|\theta)\ p(\theta)\ d\theta$$

4. The concept of informational entropy [Madankan et al., 2015], $H(Z)$, provides a mathematically rigorous framework to look for measurement acquisition parameters, $\vec{k}$, with the high information content of the reconstruction. Given a probability space $(\Omega, \mathcal{F}, p)$ (probability maps from the sigma-algebra of possible events $p : \mathcal{F} \to [0, 1]$ sigma-algebra, $\mathcal{F}$, defined on set of 'outcomes' Ω [Durrett, 2010]), we will define information of an event as proportional to the inverse probability.

$$\text{information} \equiv \frac{1}{p(z)}$$

   Intuitively, when a low probability event occurs this provides high information. The informational entropy is an *average* of the information content for a sigma algebra of events $\mathcal{F}$

$$H(Z) = \int_Z p(z) \ln \frac{1}{p(z)}\ dz \qquad p(z) = \int_\theta p(z|\theta)\ p(\theta)\ d\theta$$

Hence this entropy measure is an average of the information content for a given set of events, $\mathcal{F}$, and is proportional to the variance or uncertainty in which the set of events occur. This agrees with thermodynamic entropy; if the information containing events are completely spread out such as in a uniform distribution, the entropy is maxmized. The entropy is zero for a probability distribution in which only one event occurs. Zero information is gained when the same event always occurs ($0 \ln \frac{1}{0} = 0$). Intuitively, we want to find acquisition parameters, $\vec{k}$, for which the measurements are most uncertain

$$\max_k H(Z) \quad \Leftrightarrow \quad \min_k \int_Z dz \underbrace{\int_\theta d\theta\; p(z|\theta)\; p(\theta)}_{p(z)} \underbrace{ln \left(\int_\theta d\theta\; p(z|\theta)\; p(\theta) \right)}_{ln\; p(z)}$$

Alternatively we may consider this entropy maximization problem as a sensitivity analysis for the variance of the measurement Z, ie . $\max_k H(Z) \approx \max_k \mathrm{Var}(Z)$

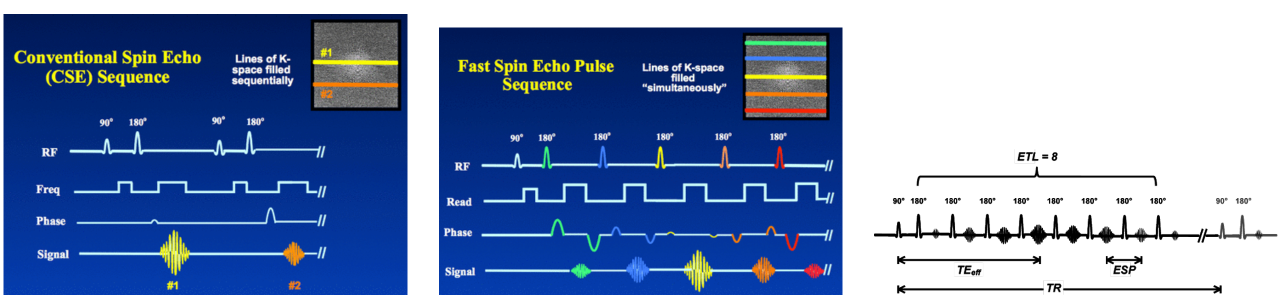
$$\begin{aligned} \bar{Z} = \mathbb{E}[Z] &= \int_Z dz\; z \underbrace{\int_\theta d\theta\; p(z|\theta)\; p(\theta)}_{p(z)} \\ \mathbb{E}[(Z-\bar{Z})^2] &= \int_Z dz\; (z-\bar{Z})^2 \underbrace{\int_\theta d\theta\; p(z|\theta)\; p(\theta)}_{p(z)} \\ &\propto \int_Z dz\; (z-\bar{z})^2 \int_\theta d\theta \exp\left(-\frac{1}{2}\|z-\mathcal{G}(\vec{k},\theta)\|^2_{\Sigma_z}\right) \exp\left(-\frac{1}{2}\|\theta-\hat{\theta}\|^2_{\Sigma_\theta}\right) \end{aligned}$$

Probilistic integrals may be computed from uncertainty quantification techniques [Fahrenholtz et al., 2013].

D WIP - Echo train length

In conventional spin-echo imaging, two basic timing parameters are required, repetition time (TR) and echo time (TE), Figure 5(a). Similar to fast spin echo (FSE) imaging, the acquistion is setup to acquire multiple lines of k-space in a single TR. In this situation, TE is replaced by effective echo time and addition parameters are needed:

- $\mathrm{TE}_{\mathrm{eff}} \equiv$ the time at which the central lines of k-space are being filled.
- Number of echoes $\equiv$ called echo train length (ETL)
- Time between echoes $\equiv$ called echo spacing (ESP)



(a) (b) (c)

Figure 5: (a)

TODO - need to update signal model for multiple read out lines