The information extraction methods for terahertz spectra-in the time and frequency domains

Reporter: ZHANG, Hongzhen

A simple summary:

- This is my research project during my master's program at Tianjin University.
- This project was launched at March 2018, and ended at December 2019.
- My degree thesis on this research was defensed at May 2020.
- Our research group includes a master's student (me), a professor (my supervisor), and two undergraduate students.
- This research is funded by National Natural Science Foundation of China (NSFC) (Grant No.61675151).

Here are some outcomes:

1) Extension of Terahertz Time Domain Spectroscopy: A Micron-level Thickness Gauging Technology
   Hongzhen Zhang, Lili Shi, Mingxia He*. *Optics Communications*, 506 (2022) 127597.
2) Terahertz Thickness Measurement Based on Stochastic Optimization Algorithm (In Chinese)
   Hongzhen Zhang, Mingxia He*, Lili Shi, Pengfei Wang. *Spectroscopy and Spectral Analysis*, 40(2020) 3066-3070.

3) A terahertz non-polar material detection technology based on Rouard's Method
   with Mingxia He, Lili Shi and Pu Wang. *Invention patent, Patent No. CN201910303091.9*, Waiting for granting.
4) A thickness measurement technology developed with terahertz spectrum
   with Mingxia He, Lili Shi and Pu Wang. *Invention patent, Patent No. CN201811197783.1*, Granted.

CONTENTS

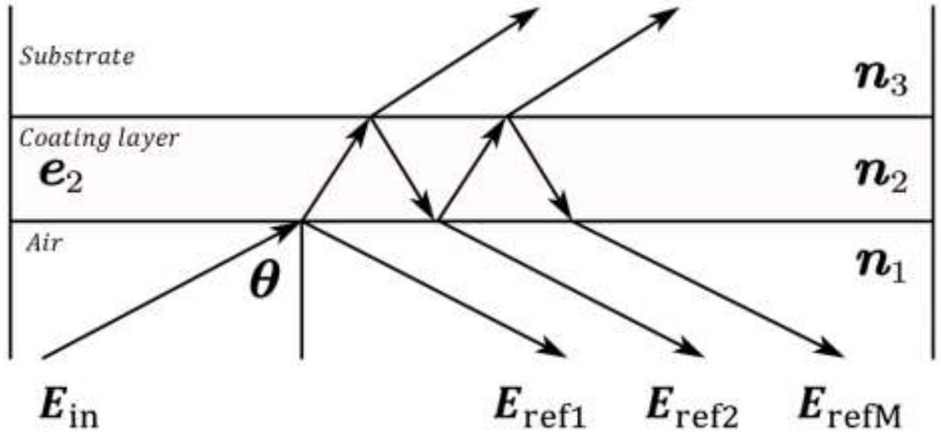
PART 1
Backgrounds

PART 2
Methods

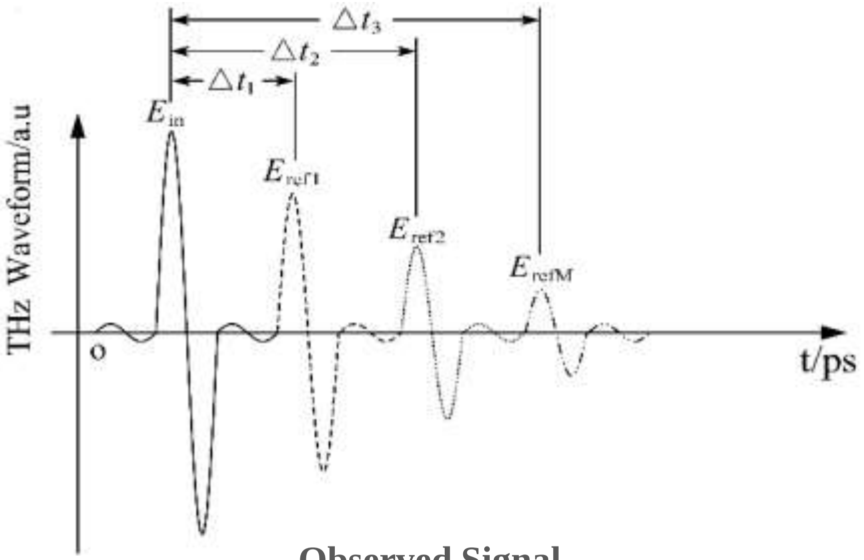
PART 3
Results and Applications

PART 4
Discussion and future directions

Part 1: Backgrounds



Coating Sample



Observed Signal

Introduction – Through an example of terahertz non-destructive testing (THz NDT)

- We would like to estimate the thickness e_2 of coating layers with THz time domain spectroscopy.
- When an incident pulse arrives at the coating surface, there will be several pulses reflected from different interfaces, such as **the first pulse** E_{ref1} from the coating surface, **the second pulse** E_{ref2} from the substrate surface, and the pulses E_{refM} are multiple-reflected between the coating surface and the substrate surface (after E_{ref1} and E_{ref2}).
- The observed signal contains E_{ref1} , E_{ref2}, and the multiple reflected pulses E_{refM}, but **these multiple reflections contribute much less than the first two major reflected pulses**.
- The thickness could be estimated by **1) calculating the difference between Time-of-flights (ToFs) Δt_1 and Δt_2 in the time domain**, or **2) extracting from the phase change between reflected pulses in the frequency domain**.
- This is an inverse problem, as the target is to extract the unknown information given the observed terahertz signals.

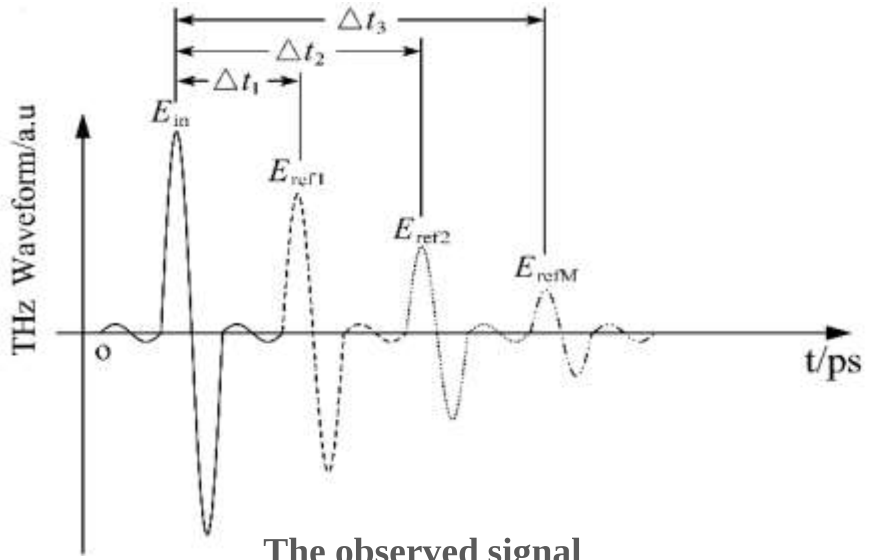
Part 2-1: Time domain method: Direct modelling with ToFs embedded

Mathematical Model:

The model here is constructed by taking all the reflected pulses contained in the observed signal into consideration, and the ToF of each pulse is directly embedded as model parameters. Based on this idea, the model could be described as:

$$E_{\mathrm{fit}}(t) = \sum_{\mathrm{i}=1,2,3\dots} E_{\mathrm{ref\,i}}(t) = \sum_{\mathrm{i}=1,2,3\dots} k_{\mathrm{i}} E_{\mathrm{in}}(t+\Delta t_{\mathrm{i}})$$

Thickness: $e_2 = \dfrac{c(\Delta t_2 - \Delta t_1)}{2n_2}$



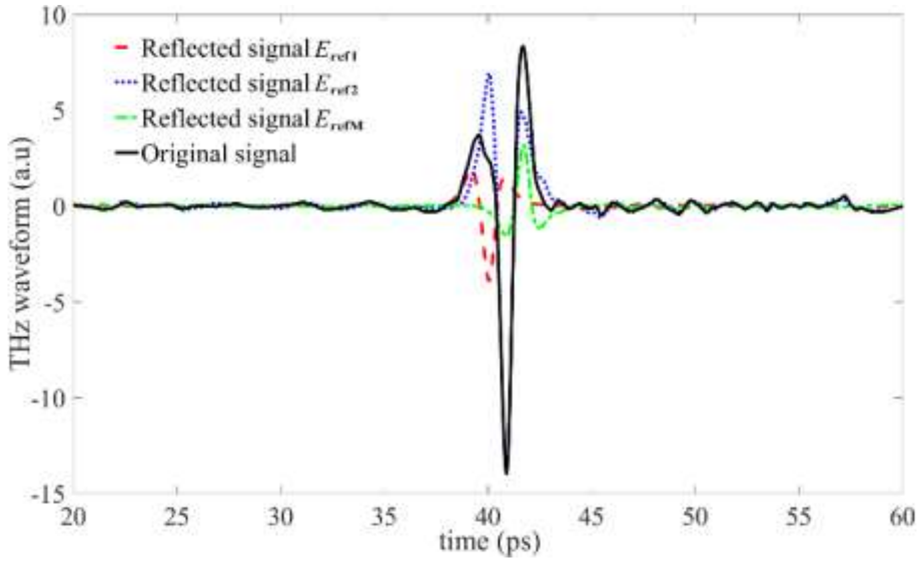
The observed signal

Loss function:

As aforementioned, the multiple reflected pulses (reflected after E_{ref1} and E_{ref2}) contribute much less to the observed signal, therefore, to avoid overfitting, we added penalty on their coefficients (in other words, we would like to only extract the ToFs of E_{ref1} and E_{ref2}):

$$Loss = \sum_{k=1}^{N} (E_{\mathrm{fit}}(t_k) - E_{\mathrm{mea}}(t_k))^2 + \lambda \cdot \sum |k_j|, j > 2$$

$E_{\mathrm{mea}}(t_k)$ is the observed signal.



The decomposed signal

Optimizer:

The incident pulse in our case does not has an analytical expression, therefore we could hardly calculate the gradient of the above loss function with respect to the unknown parameters Δt_{i}.

Metaheuristic optimizing algorithms make few or no assumptions about the problem being optimized, and a differentiable target function is not required. In our case, we took advantage of Genetic Algorithm and Differential Evolution Algorithm to calibrate model parameters (**minimize the loss function**).

Part 2-2: Frequency domain method: Based on transfer function and refractive index modelling

Mathematical Model:

This model is constructed based on the transfer function of the medium layer. The thickness could be extracted from the phase change $\varphi = \frac{2\pi}{\lambda} n_2 e_2$. The model could be described as:

$$H(\omega) = r_{1-2} + \frac{t_{1-2} r_{2-3} t_{2-1} exp(-2i\frac{2\pi}{\lambda} n_2 e_2)}{1 - r_{2-1} r_{2-3} exp(-2i\frac{2\pi}{\lambda} n_2 e_2)}$$

The refractive index varies at different frequencies, to model this relationship, we leveraged the dielectric model to parameterize the refractive index of coating layers:

$n(\omega) = \sqrt{\varepsilon(\omega)}$ Debye model (a dielectric model): $\varepsilon(\omega) = \varepsilon_\infty + \frac{\varepsilon_1 - \varepsilon_\infty}{1 + i\omega\tau_1} + \sum_{l=2}^{m} \frac{\varepsilon_l - \varepsilon_{l-1}}{1 + i\omega\tau_l}$

This frequency domain method could be easily generalized to mediums with multiple layers - *the Rouard's Method.*

Loss function:

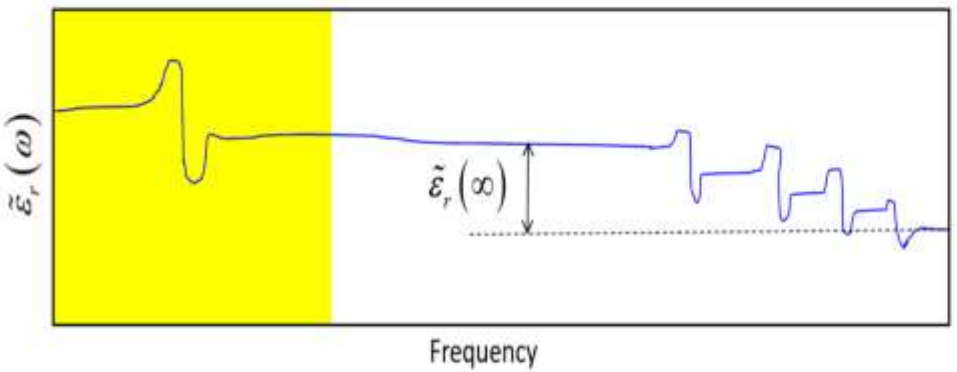
The fitted signal would be transformed back to the time domain in order to calculate the loss. The penalty term is not required in this frequency domain method as the transfer function seamlessly considers all the reflected pulses:

$$E_{\text{fit}}(t) = IFFT(E_{\text{in}}(\omega) H(\omega))$$

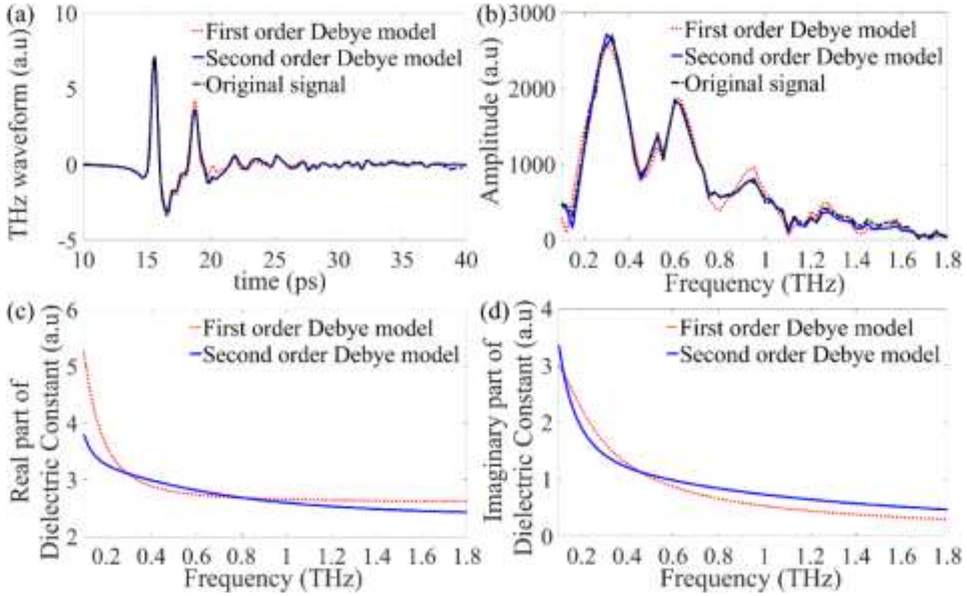
$$Loss = \sum_{k=1}^{N} (E_{\text{fit}}(t_k) - E_{\text{mea}}(t_k))^2$$

Optimizer:

The model parameters (n_2, e_2) could be calibrated by Genetic Algorithm (GA) or Differential Evolution (DE) Algorithm.

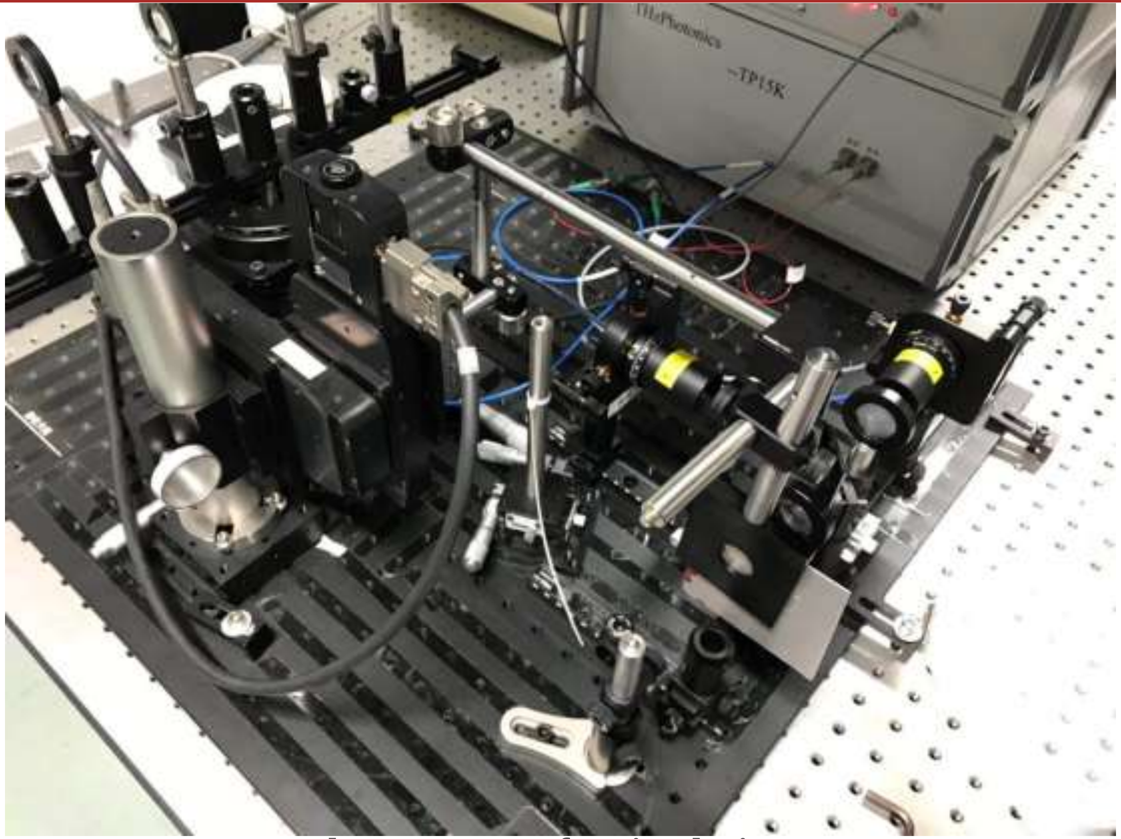


Dielectric property: Superposition of responses



Fitting performance with Debye models in different orders

Part 3: Results and Applications



The prototype of optics device

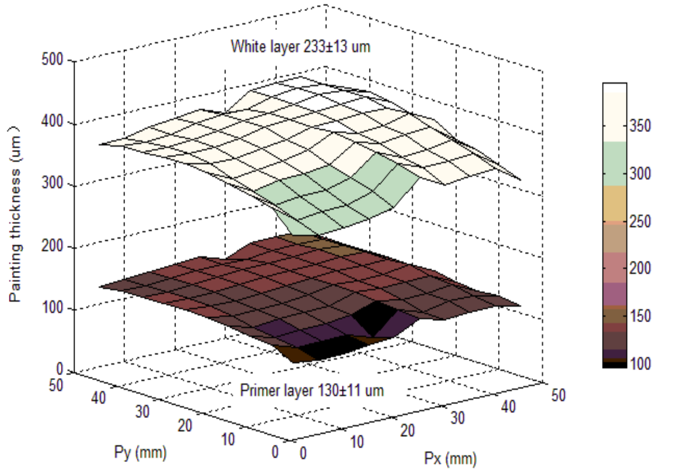
Thickness measurement results*

Coating Samples	Time domain method	Reference	Frequency domain method
Red	78.1	81.8	79.9
Base	224.3	228	228.2
Black	53.1	54.8	52.7
Pearl white	205.3	207	219.7
Golden	38.9	40.1	29.7

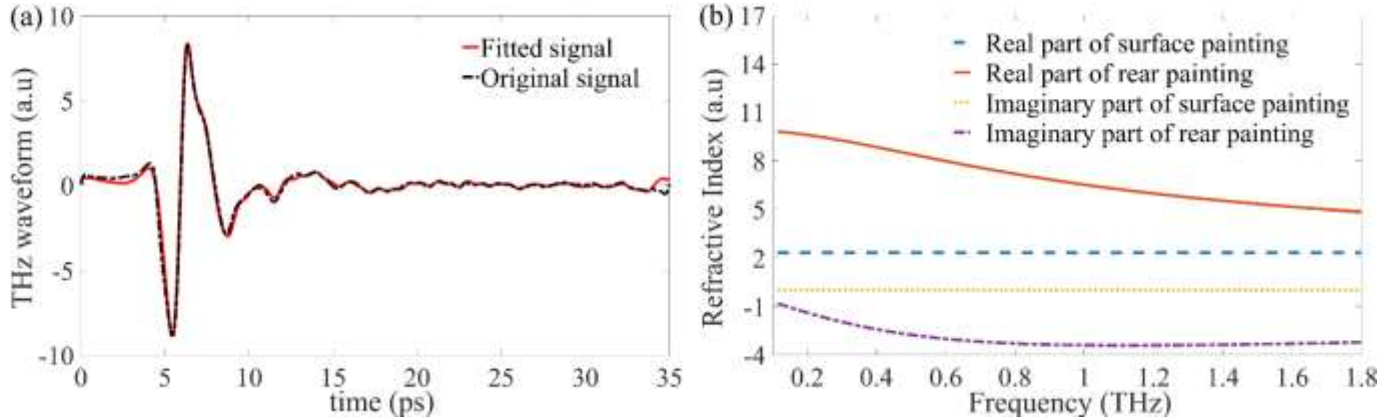
*all the results are in (μm)

Other applications:

Except for thickness measurement, our method could also be applied on some other scenarios, such as the ToF-based imaging and calibration for unknown materials. Below are some primary results delivered by our research group.



Imaging based on ToF



Material calibration for a double-layer coating sample

Part 4: Discussion: Future working directions

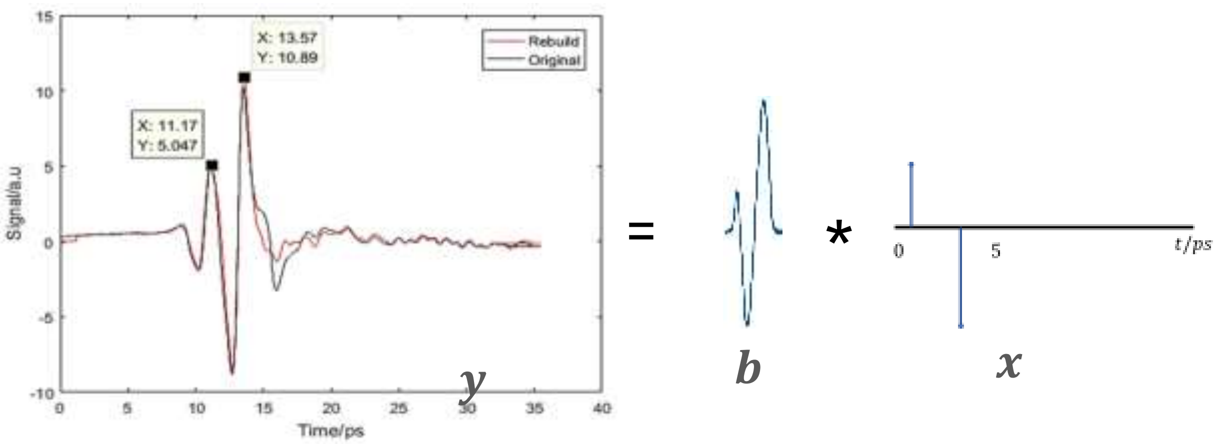
Two significant limits of current information extraction (spectra analysis) methods:

Although our methods perform quite well, there all still two significant limits due to the optimizing algorithms and the mathematical models.

- Metaheuristic optimizing algorithms require a huge population as well as a large number of iterations to converge at an optimal solution, which will significantly slow down the processing time of our methods.
- Due to the complicated form of our mathematical models and the high dimension of parameter space (when generalize the frequency domain method to mediums with 4 layers, the number of unknown parameters will be more than 20), the loss function is usually not convex. It demonstrates lots of local optimal solutions.

Future direction - Short and Sparse Deconvolution (SaSD) method:

We have done some primary works on SaSD method.



$$\min_{\boldsymbol{b},\boldsymbol{x}} \frac{1}{2}||\boldsymbol{y}-\boldsymbol{b}*\boldsymbol{x}||_2^2+\lambda||\boldsymbol{x}||_1$$

Reformulate

$$\min_{\boldsymbol{b},\boldsymbol{x}} \frac{1}{2}||\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}||_2^2+\lambda||\boldsymbol{x}||_1$$

$$\boldsymbol{A}=\begin{bmatrix} i_0 & i_n & \cdots & i_1 \\ i_1 & i_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & i_n \\ i_n & i_{n-1} & \cdots & i_0 \end{bmatrix}$$

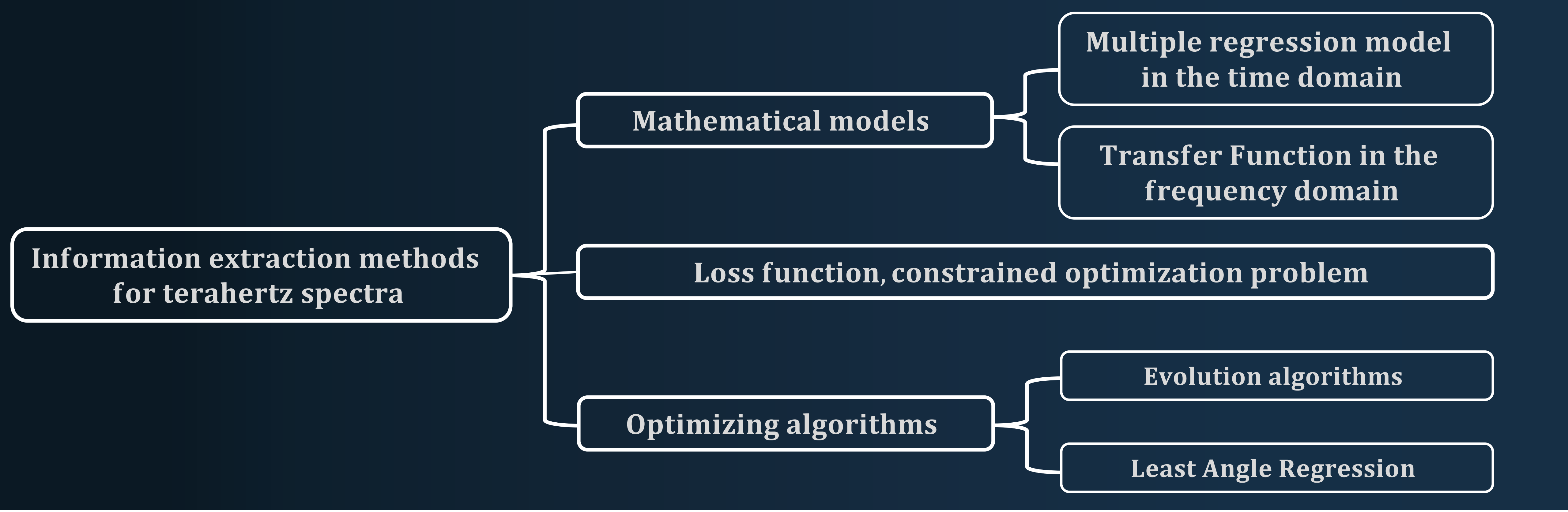
$$\boldsymbol{b}=[i_0,i_1,i_2,\ldots,i_{n-1},i_n]$$

1) The original signal $\boldsymbol{y}$ could be described as a convolution between a basic kernel $\boldsymbol{b}$ (in our case, it is the input pulse) and a sparse vector $\boldsymbol{x}$.
2) This convolution could be reformulated into a Linear System with an observation matrix $\boldsymbol{A}$ that is constructed by rolling the basic kernel $\boldsymbol{b}$.
3) When $\boldsymbol{b}$ is known as prior information, the above problem could be solved by general LASSO algorithms (Here the LARS algorithm is applied).

This method is much faster, however, it is very sensitive to signal noise -*More robust*.

Appendix

Appendix 1: The whole framework of our methods



Models

- Multiple Regression Model

$$E_{\text{fit}}(t) = \sum_{\text{i}=1,2,3\ldots} E_{\text{ref i}}(t) = \sum_{\text{i}=1,2,3\ldots} k_{\text{i}} E_{\text{in}}(t + \Delta t_{\text{i}})$$

- Model based on transfer function

$$E_{\text{fit}}(t) = IFFT(E_{\text{in}}(\omega)H(\omega))$$

Loss function

- Sum of squared residuals

$$y = \sum_{k=1}^{N} (E_{\text{fit}}(t_k) - E_{\text{mea}}(t_k))^2$$

- Constrained by L1 norm

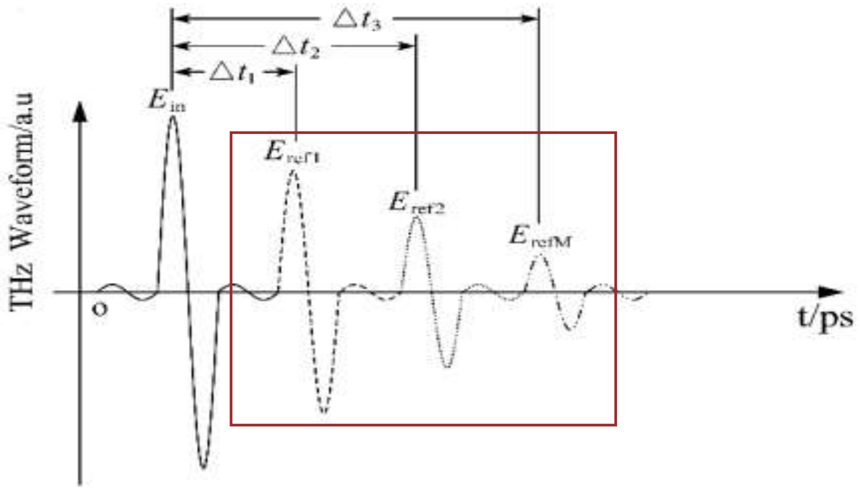
$$\lambda \cdot \sum |k_j|, j > 2$$

Optimizing Algorithms

- Make few or no assumptions about the problem being optimized.
- For multidimensional real-valued functions but a differentiable target function is not required.

Appendix 2: Details on the time domain method

The multiple regression model:



- E_{ref1} and E_{ref2} constitute the majority of the observed signal:

$$E_{major}(t) = E_{ref\,1}(t) + E_{ref\,2}(t) = k_1 E_{in}(t + \Delta t_1) + k_2 E_{in}(t + \Delta t_2)$$

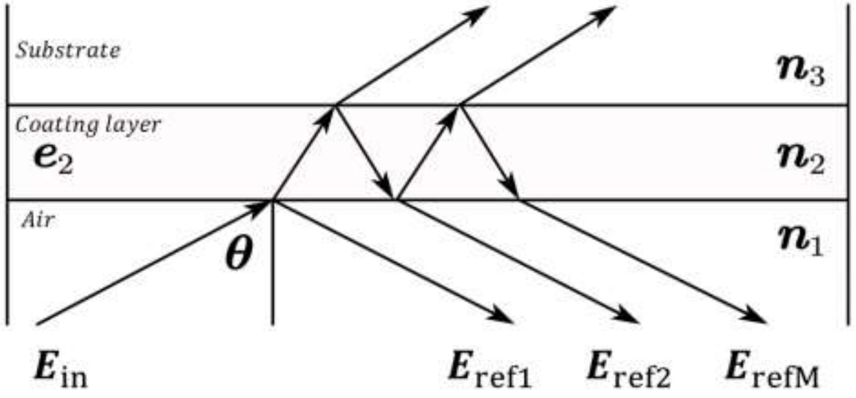
- The time difference between two adjacent multiple reflected pulses is the same:

$$E_{refM} = \sum_{j=1}^{\infty} (-1)^j M_j E_{in}\big(t + \Delta t_2 + j(\Delta t_2 - \Delta t_1)\big)$$

- The overall signal consists of E_{ref1} and E_{ref2} and multiple reflected pulses:

$$E_{total}(t) = E_{major} + E_{refM}$$

How the thickness is calculated:



- According to the Time-of-flight theory:

$$\Delta l:\ \Delta l = 2n_2 e_2 = c\Delta T,\ \Delta T = \Delta t_2 - \Delta t_1$$

- The normalized refractive index could be solved by Fresnel Equation:

$$k_1 = r_{1,2} = \frac{n_2 - n_1}{n_2 + n_1}\ (Fresnel\ Equation)$$

- The thickness of medium could be solved inversely:

$$e_2 = \frac{c\Delta T}{2n_2} = \frac{c(\Delta t_2 - \Delta t_1)}{2n_2}$$

Appendix 3: Physics behind the frequency domain method

How to get the transfer function of coating mediums?

Transform the multiple regression model in the time domain into the frequency domain with Fourier Transform, we would get a geometric series:

$$E_{\text{total}}(t) = E_{\text{major}} + E_{\text{refM}}$$

Fourier Transform

$$H(\omega) = r_{1-2} + \sum_{k=1}^{\infty} t_{1-2}\left[r_{2-3} exp\left(-2in_2 \frac{2\pi}{\lambda} e_2\right)\right]^k t_{2-1}$$

Geometric series

$$H(\omega) = r_{1-2} + \frac{t_{1-2} r_{2-3} t_{2-1} exp\left(-2in_2 \frac{2\pi}{\lambda} e_2\right)}{1 - r_{2-1} r_{2-3} exp\left(-2in_2 \frac{2\pi}{\lambda} e_2\right)}$$

Summarize all these terms with equal ratio, we could get the transfer function.

The relation between refractive index and permittivity

Refractive index is actually a function of frequency, in the time domain, we treat it as a constant (normalized), however, in the frequency domain, we would like to recover this function relationship:

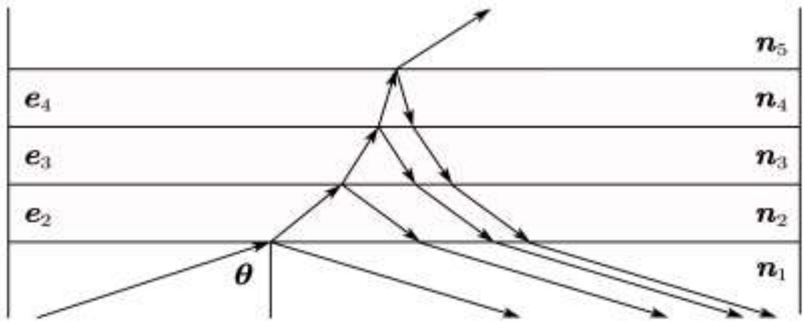
$$\tilde{n} = \mathrm{g}(\omega)$$

Refractive index is the square root of permittivity (magnetic conductivity is 1):

$$\tilde{n} = n + j\kappa = \pm\sqrt{\tilde{\varepsilon_r}}$$

Therefore we could use some dielectric models to describe the function relationship between refractive index and frequency. Theoretically speaking, we could even use a polynomial model to parameterize the refractive index under varying frequency.

Details on Rouard's Method



A multi-layer medium

This is a recursive idea, starting from a medium sample with single layer, its transfer function could be calculated, and the whole transfer function would be treated as a normalized reflection coefficient and then be used to replace the reflection coefficient between this layer and the last layer:

Single layer: $\boxed{r_4} = \dfrac{r_{3-4} + r_{4-5} e^{-2j\varphi_4}}{1 + r_{3-4} r_{4-5} e^{-2j\varphi_4}}$

Replace $\boldsymbol{r_{3-4}}$

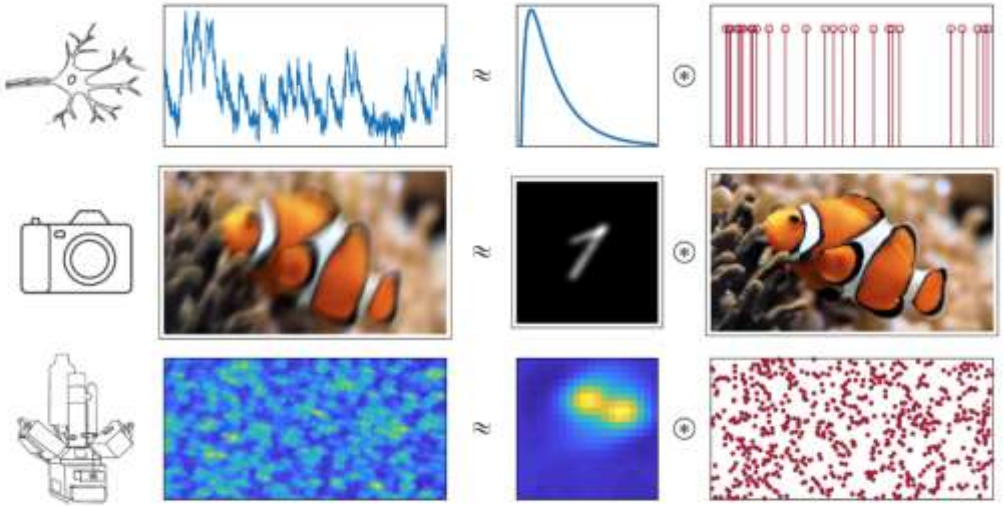
Double layer: $\boxed{r_3} = \dfrac{r_{2-3} + r_4 e^{-2j\varphi_3}}{1 + r_{2-3} r_4 e^{-2j\varphi_3}}$

Replace $\boldsymbol{r_{2-3}}$

Triple layer: $\boxed{r_2} = \dfrac{r_{1-2} + r_3 e^{-2j\varphi_2}}{1 + r_{1-2} r_3 e^{-2j\varphi_2}}$

We could generalize this model to mediums with any number of layers, however, the number of unknow parameters would increase dramatically and it would be challenging to get an accurate estimation.

Appendix 4: Details on Sparse Deconvolution and LARS algorithm



Calcium Imaging, Image Deblurring, Scanning tunneling microscopy

Basic Idea of Sparse Deconvolution

Signals (in medical/scientific/natural imaging) can often be modeled as the convolution between a basic, recurring motif $\boldsymbol{b}$ and a sparse vector $\boldsymbol{x}$. By optimizing the function below we could find out the best kernel and sparse vector:

$$\min_{b,x} \frac{1}{2}||\boldsymbol{y} - \boldsymbol{b} * \boldsymbol{x}||_2^2 + \lambda||\boldsymbol{x}||_1$$

which could be reformulated into a Linear System with an observation matrix $\boldsymbol{A}$:

$$\min_{b,x} \frac{1}{2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{x}||_1$$

where matrix $\boldsymbol{A}$ is constructed by rolling kernel b. When b is known as prior, the above problem could be solved by LASSO algorithms like LARS and Coordinate Descent.

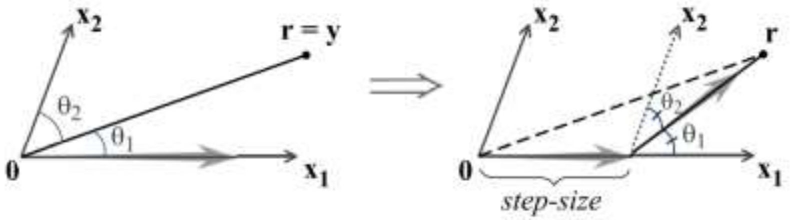
Matrix $\boldsymbol{A}$ and Kernel $\boldsymbol{b}$ in our case:

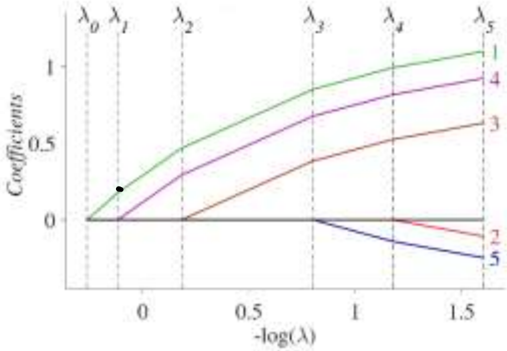
$$\boldsymbol{A} = \begin{bmatrix} i_0 & i_n & \cdots & i_1 \\ i_1 & i_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & i_n \\ i_n & i_{n-1} & \cdots & i_0 \end{bmatrix}$$

$$\boldsymbol{b} = [i_0, i_1, i_2, \dots, i_{n-1}, i_n]$$

Kernel $\boldsymbol{b}$ is the discretized input pulse, and matrix $\boldsymbol{A}$ is designed by moving kernel $\boldsymbol{b}$ in different lags.

Details on LARS (Least Angle Regression) algorithm*





Starting from all zeros, LARS picks the predictor $\boldsymbol{x}_1$ that makes least angle (i.e., $\theta_1 < \theta_2$) with the current residual $\boldsymbol{r}$ and moves in its direction until $\theta_1 = \theta_2$ where LARS picks $\boldsymbol{x}_2$ and changes direction (based on the KKT condition).

*Tabassuma M. N. and Ollila E. (2017)

Appendix 5: Differential Evolution Algorithm

Algorithm 1: DE algorithm

Input: Population: M; Dimension: D; Genetation: T
Output: The best vector (solution) - Δ

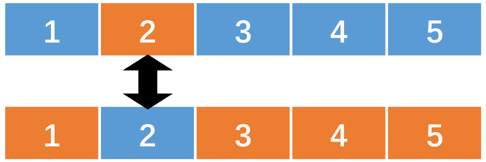
$t \leftarrow 1(initialization)$;
for $i = 1$ *to* M **do**
  for $j = 1$ *to* D **do**
    $x^j_{i,t} = x^j_{min} + rand(0,1) \cdot (x^j_{max} - x^j_{min})$;
  end
end
while $(|f(\Delta)| \geq \varepsilon)$ *or* $(t \leq T)$ **do**
  for $i = 1$ *to* M **do**
    ▶ *(Mutation and Crossover)*
    for $j = 1$ *to* D **do**
      $v^j_{i,t} = Mutation(x^j_{i,t})$;
      $u^j_{i,t} = Crossover(x^j_{i,t}, v^j_{i,t})$;
    end
    ▶ *(Greedy Selection)*
    if $f(\boldsymbol{u}_{i,t}) < f(\boldsymbol{x}_{i,t})$ **then**
      $\mathbf{x}_{i,t} \leftarrow \mathbf{u}_{i,t}$;
      if $f(\boldsymbol{x}_{i,t}) < f(\Delta)$ **then**
        $\Delta \leftarrow \mathbf{x}_{i,t}$;
      end
    else
      $\mathbf{x}_{i,t} \leftarrow \mathbf{x}_{i,t}$;
    end
  end
  $t \leftarrow t + 1$;
end
return the best vector Δ;

Mutation

- Take difference on random selected parent individuals to generate mutation vector.
- Generate the cross population by add mutation vectors to the best individual:
  $M_i = parent_i - parent_j;\ C_i = Individual_{best} + F_i \times M_i$
- Test the feasibility (within bounds), if not, perform the mutation operation again.

Cross

- Take the parent individuals as the parent population.
- Perform the cross operation on the cross population and the parent population.
- Generate the candidate child population.
- Test the feasibility (within bounds), if not, perform the cross operation again.



Selection

- Select the better individual by comparing the parent and candidate child individuals.
- Replace the individual in the parent population and generate the next generation population.