Dynamic, Data-Driven Modeling of Nanoparticle Self-Assembly Processes

Yu Ding[1], Jianhua Huang[1], Chiwoo Park[2], Tao Liu[2] and Chuck Zhang[3]

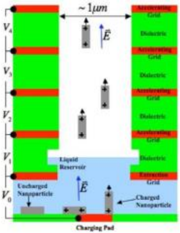
[1] Texas A&M University
[2] Florida State University
[3] Georgia Institute of Technology

Sponsored by AFOSR DDDAS program (FA9550-13-1-0075)
and DURIP program (FA9550-16-1-0110)

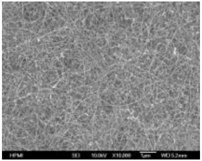
Nanoparticles

- Nanoparticles are ultrafine particles characterized by their nanoscale dimension ranging from 1 to 100 nanometers.
- Their unique properties have been studied and exploited for many defense and security related applications.

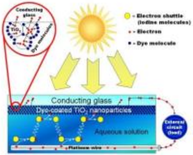
(a) Nanoparticles as propellants of satellites and space craft propulsion (Louis et. al. 2007)



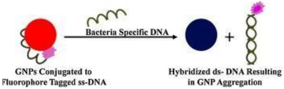
(b) Nanocomposites with excellent mechanical and electric properties (FSU HPMI)



(c) Photovoltaic catalyst for solar cell (NSF)



(d) Sensing toxic biological weapons (Venkata et. al. 2011)



- These applications require nanoparticle products of precisely controlled sizes and shapes, because the functionalities of the nanoparticles are determined by their sizes and shapes.

Statement of Research

- Due to the random nature of a chemical synthesis process of nanoparticles (colloidal self-assembly), nanoparticles produced therein are prone to having broad distributions in size and shape.
- Producing nanoparticles with concentrated size and shape distributions is a long-time desire of nano scientists.
- For controlling the process for better size and shape concentration, one should have a good process model to describe size and shape changes in the self-assembly process.
- However, classical theory of particle crystallization does not appear to exactly describe the self-assembly system. Hard to find the first-principle-based theory for explaining the full scale nanoparticle self-assembly.

Dynamic Data-Driven Approach: Our research objective is to build a predictive model of particle size and shape changes, i.e., P(Size_t , Shape_t), using multiscale process measurements.

Combining multiscale metrologies

There are multiple measurement techniques that can measure particle sizes and shapes, but there is no single technique that provides sufficient information to build the predictive model.

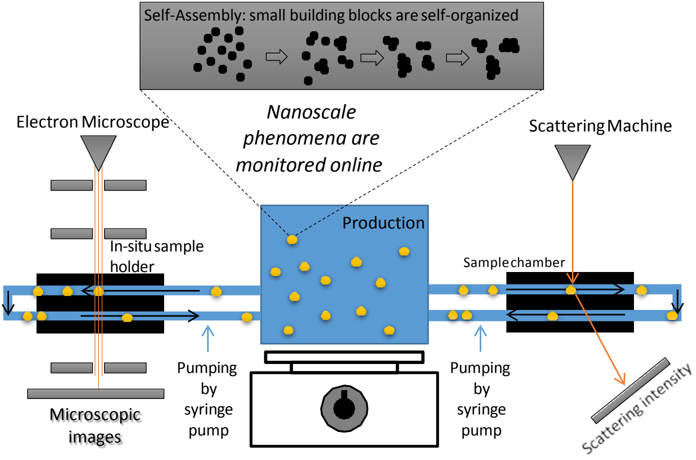
- Dynamic light scattering (DLS) provides particle size information conditioned on that shape information is known; i.e., DLS can only provide $P(\text{Size}_t \mid \text{Shape}_t)$.
- Electron microscope (TEM) can provide both size and shape information but do so for a very small portion of sample. The output may not be statistically reliable. However, the shape distribution, $P(\text{Shape}_t)$, is still estimable with the small portion since the variety of shapes is relative small.

Combining two instruments can provide the full picture,

$$P(\text{Size}_t \mid \text{Shape}_t) P(\text{Shape}_t) = P(\text{Size}_t, \text{Shape}_t).$$

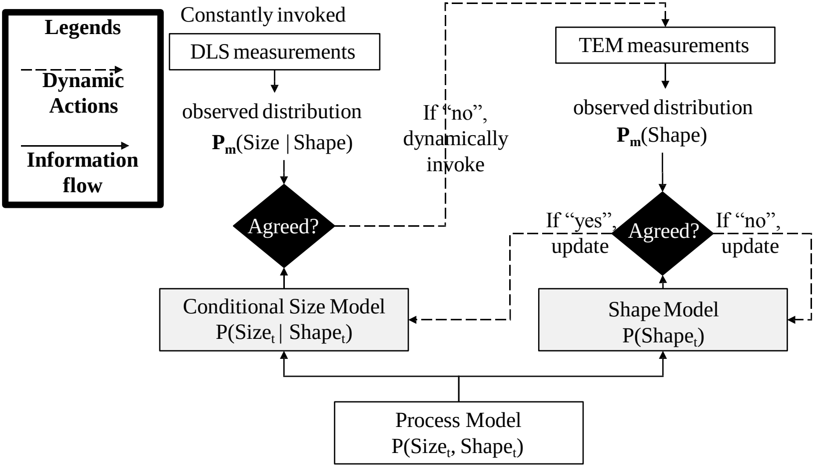
Combining multiscale metrologies

We will use both TEM and DLS measurements to build a predictive model $P(\text{Size}_t\,,\,\text{Shape}_t)$. The most desirable is taking full information out of the two instruments while minimizing expensive TEM operations.



Overall DDDAS Strategy

We expect that $P(Shape_t)$ changes slower than $P(Size_t \,|Shape_t)$! Trigger TEM operations only when potential changes in $P(Shape_t)$ are detected. But, how?



My research: Big data challenge in realtime monitoring

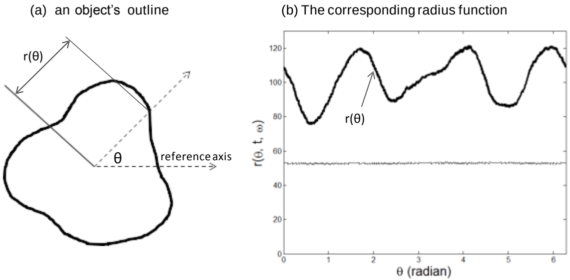
To implement the strategy, we need to solve the following research problems:

- Task 1. Data Analysis: DLS for Particle Sizes (Li et al., 2015), TEM for Particle Sizes and Shapes (Park et al., 2013; Konomi et al., 2013; Qian et al., 2015; Agbabiaka et al., 2013)
- Task 2. Dynamic Process Modeling: How do we describe the process of changes in particle size and shape distributions? (Park, 2014; Woehl et al., 2013)
- Task 3. Dynamic Model Update: How do we update the process model with either of DLS or TEM data? (Park, 2014)
- Task 4. Dynamic Sensor Invocation: When do we need to invoke TEM operations? (Park and Shrivastava, 2014; Qian et al., 2015)

Process modeling

We represent 'size and shape' of a nanoparticle as the outline of the nanoparticle (Park, 2014, Technometrics).

- $r(\theta)$ represents the outline, which quantifies the distance from the center to the outline along angle θ.



- $r(\theta, t)$ represents the time change of the outline by a nanoparticle self-assembly process, and

$$Size_t = \frac{1}{2\pi}\int r(\theta,t)d\theta, \quad Shape_t = \frac{1}{Size_t} r(\theta,t).$$

Process modeling

The process can be represented in a parametric form,

$$r(\theta, t) = \sum_{m=1}^{M} \sum_{n=1}^{N} \alpha_{m,n}\phi_m(t)\gamma_n(\theta), t \geq 0 \text{ and } \theta \in [0, 2\pi), \tag{1}$$

- $\varphi_m(t)$'s and $\gamma_n(\theta)$'s are spline basis functions.
- $\alpha_{m,n}$ are degrees of freedom (basis coefficients).

$\alpha = (\alpha_{m,n}) \sim P(\alpha)$ quantifies the randomness on the outline change in a self-assembly process, and $P(Size_t, Shape_t)$ can be fully characterized by

$$P(Size_t, Shape_t) := P(r(t, \theta)) = \int P(r(t, \theta)|\alpha) dP(\alpha) \tag{2}$$

Dynamic Model Updates

The prior distribution $P(\boldsymbol{\alpha})$ is updated with new data, from either DLS or TEM.

$$P(\boldsymbol{\alpha}|\text{New Data}) \propto P(\text{New Data}|\boldsymbol{\alpha})P(\boldsymbol{\alpha})$$

The likelihood model $P(\text{New Data}|\boldsymbol{\alpha})$ is different depending on the type of data. Note that the DLS data only carries size information,

$$\begin{aligned} Size_t &= \frac{1}{2\pi}\int r(\theta, t)d\theta \\ &= \frac{1}{2\pi}\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{m,n}\phi_m(t)\int \gamma_n(\theta)d\theta \\ &= \sum_{m=1}^{M}\left(\frac{1}{2\pi}\sum_{n}\alpha_{m,n}\right)\phi_m(t). \end{aligned} \tag{3}$$

We formulate the two likelihood models

- For DLS:a link function relating size observations to $\alpha_m = \frac{1}{2\pi}\sum_n \alpha_{m,n}$.
- For TEM:a link function relating shape observations to individual $\alpha_{m,n}$'s.

Dynamic Model Updates

- Exact block Gibbs sampler for the model updatewith TEM likelihood model can be found in our technometrics paper (Park, 2014), but it is also applicable for the model update with the DLS likelihood model.
- It is based on the Gibbs samplers proposed for an infinite mixture model (Papas 2008) and is modified for the mixture of truncated distributions.
- For more computationally efficient update, we are working on anactive set type optimizer for updating the infinite mixture model.

Triggering TEM Operation

A remaining problem is whether to trigger TEM operation at time $t+1$. Following our paper (Li et al., 2015), we use the general likelihood ratio test with DLS data. Conceptually,

- Collect new DLS Data at time $t+1$.
- Conditioned on that the current shape model $P(Shape_t)$ remains true in time $t+1$, predict $P(Size_{t+1}|Shape_t)$ from the predictive model, and obtain the joint distribution $P(Size_{t+1}|Shape_t)P(Shape_t)$.
- Compute the likelihood of observing the DLS data given the particle sizes and shapes are distributed as $P(Size_{t+1}|Shape_t)P(Shape_t)$.

If the likelihood is lower than a certain threshold, either the current shape model $P(Shape_t)$ or the model prediction $P(Size_{t+1}|Shape_t)$ is wrong, so we need to trigger new TEM to figure out the reason.

Simultaion Study: TEM Triggering

We simulated data at process times $t \in \{1, ..., 50\}$ from the following noisy process

$$r(t,\theta) = \sum_{m=1}^{M}\sum_{n=1}^{N} \alpha_{m,n}\phi_m(t)\gamma_n(\theta) + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}),$$

where

- Assumed Distribution. $\boldsymbol{\alpha} \sim \omega_1 \mathcal{N}_Q(\boldsymbol{\mu}_1, \sigma_0^2 \boldsymbol{I}) + \omega_2 \mathcal{N}_Q(\boldsymbol{\mu}_2, \sigma_0^2 \boldsymbol{I})$.
- Parameters. $\Theta = \{\omega_1, \omega_2, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \sigma_0^2, \sigma^2\}$

If the data to be simulated is from

- Base process: We use the pre-determined values for $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and set $\sigma_0^2 = 1$, $\sigma^2 = 1$, $\omega_1 = 0.7$ and $\omega_2 = 0.3$.
- Size Change Scenario: We changed $\boldsymbol{\mu}_k$ to $\boldsymbol{\mu}_k + \delta\sigma_0 \mathbf{1}$.
- Shape Change Scenario: We changed $\boldsymbol{\mu}_k$ to $\boldsymbol{\mu}_k \rightarrow \boldsymbol{\mu}_k + \delta\sigma_0 \boldsymbol{e}$.

Simultaion Study: TEM Triggering

The ARL performance is averaged over 500 replicated cases (ARL = average time to detect changes / measurement rate; smaller is better).

Table : ARL performance for size changes ($\boldsymbol{\mu}_k \rightarrow \boldsymbol{\mu}_k + \delta\sigma_0\mathbf{1}$)

v	Magnitude of Change (δ)						
	0.0	0.5	1.0	1.5	2.0	2.5	3.0
0.05	21.74	1.25	1.02	1.00	1.00	1.00	1.00
0.01	166.67	1.31	1.04	1.00	1.00	1.00	1.00
0.0027	$>$ 500.00	1.35	1.06	1.00	1.00	1.00	1.00

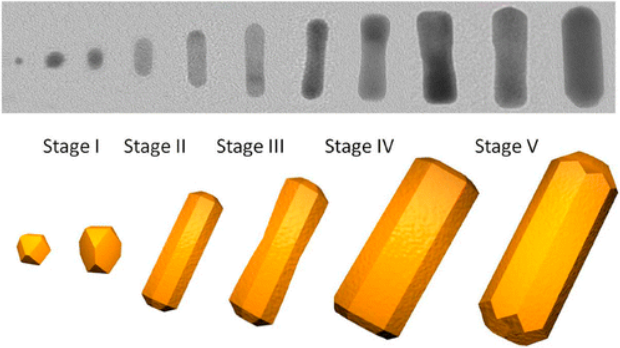
Table : ARL performance for shape changes ($\boldsymbol{\mu}_1 \rightarrow \boldsymbol{\mu}_1 + \delta\sigma_0\boldsymbol{e}$)

v	Magnitude of Change (δ)						
	0.0	0.5	1.0	1.5	2.0	2.5	3.0
0.05	21.74	13.51	3.03	1.55	1.09	1.01	1.00
0.01	166.67	50.00	3.87	1.87	1.44	1.16	1.09
0.0027	$>$ 500	71.42	4.03	1.88	1.44	1.16	1.10

Real Data Study: TEM Triggering

Summary of experimental set-up:

- Growth process: we implemented the seed-mediated nanoparticle growth developed by Nikoobakht (2003).

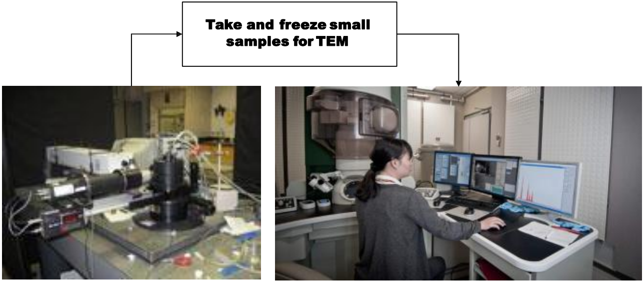


* Source: **Growth Mechanism of Gold Nanorods, Park et. al.**, *Chemistry of Materials* **2013** *25* (4), 555

Real Data Study: TEM Triggering

Summary of experimental set-up:

- The whole implementation was performed in the DLS machine for continuously taking DLS measurements.



BI-200SM (DLS/SLS)

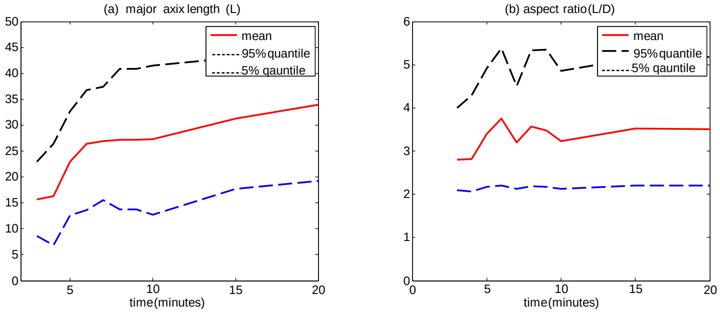
JEM-200F (spatial resolution: 0.11 nm)

- Samples were regularly taken from the DLS machine to prepare TEM samples, which was manually done since we do not have in-situ TEM yet; we will get an in-situ TEM later this year with our recent DURIP grant and can automate this measurement process.

Real Data Study: TEM Triggering

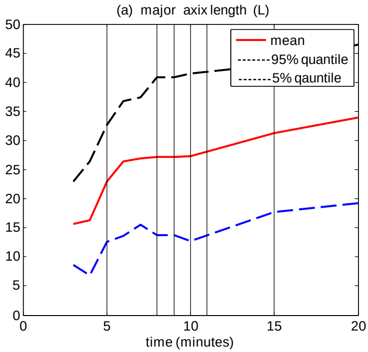
Data Description:

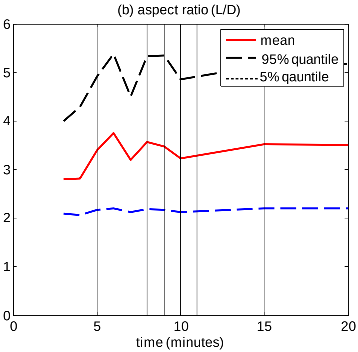
- 96 DLS data and 20 TEM data were taken regularly in between 0 and 20 minutes from the start of the process.
- Both DLS and TEM data taken in between 0 and 4 minutes were used to learn the initial predictive model $r(\theta, t)$.
- Ground truths for $P(Size_t\,, Shape_t)$ can be constructed by using all of the collected data as follows:



Real Data Study: TEM Triggering

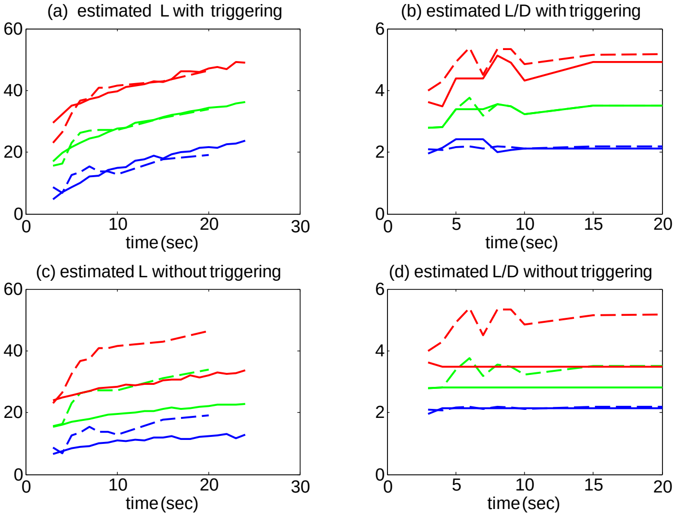
TEM triggering process was initiated after $t = 5$ minutes, following our DDDAS approach. Only 6 TEM operations are triggered in between 5 and 20 minutes.





Real Data Study: TEM Triggering

The DDDAS output was compared with two baselines, ground-truth and the naive predictive model without updates. (solid lines: estimated, dotted lines: ground truth)



Summary and On-going Works

Summary:

- **DDDAS** enables a predictive model for characterizing the evolution in particle size and shape during a self-assembly process
- **DDDAS** empowers on-line model update and adaptive TEM triggering that makes better use of expensive resources
- **DDDAS framework** leads to elimination of 9 out of 15 potential TEM operations while maintaining good predictive powers.

Future plan:

- Our next target is to make use of the dynamic modeling capability **enabled by DDDAS** to develop a model-predictive control that can potentially steer the nanoparticle self-assembly process as desired.

Agbabiaka, A., M. Wiltfong, and C. Park (2013). Small angle x-ray scattering technique for the particle size distribution of nonporous nanoparticles. *Journal of Nanoparticles 2013*.

Konomi, B., S. S. Dahvala, J. Z. Huang, S. Kundu, D. Huitink, H. Liang, Y. Ding, and B. K. Mallick (2013). Bayesian object classification of gold nanoparticles. *The Annals of Applied Statistics 7* (2), 640–668.

Li, X., P. Tran, T. Liu, and C. Park (2015). Simulation-guided regression approach for estimating the nanoparticle size distribution with dynamic light scattering data. *IIE Transactions Submitted*.

Park, C. (2014). Estimating multiple pathways of object growth using non-longitudinal image data. *Technometrics 56*(2), 186–199.

Park, C., J. Huang, J. Ji, and Y. Ding (2013). Segmentation, inference and classification of partially overlapping nanoparticles. *IEEE Transactions on Pattern Analysis and Machine Intelligence 35*(3), 669–681.

Park, C. and A. Shrivastava (2014). Multimode geometric profile monitoring with correlated image data and its application to nanoparticle self-assembly processes. *Journal of Quality Technology 46*(3), 216–233.

Qian, Y., J. Huang, and Y. Ding (2015). Identifying multi-stage nanocrystal growth using in-situ tem video data. *IIE Transactions Submitted*.

Qian, Y., J. Huang, X. Li, and Y. Ding (2015). Robust nanoparticles detection from noisy background by fusing complementary image information. *IEEE Transactions on Image Processing Submitted*.

Woehl, T. J., C. Park, J. E. Evans, I. Arslan, W. D. Ristenpart, and N. D. Browning (2013). Direct observation of aggregative nanoparticle growth: Kinetic modeling of the size distribution and growth rate. *Nano Letters 14*(1), 373–378.