

ON SENSOR TRADEOFFS, DESIGN, AND ESTIMATION IN BIOMOLECULAR SYSTEMS

KRISHAN KUMAR GOLA



DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI

JULY 2025

© **Indian Institute of Technology Delhi (IITD), New Delhi, 2025**

ON SENSOR TRADEOFFS, DESIGN, AND ESTIMATION IN BIOMOLECULAR SYSTEMS

by

KRISHAN KUMAR GOLA

Department of Electrical Engineering

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy
to the



INDIAN INSTITUTE OF TECHNOLOGY DELHI

JULY 2025

CERTIFICATE

This is to certify that the thesis entitled **On Sensor Tradeoffs, Design, and Estimation in Biomolecular Systems** submitted by **Krishan Kumar Gola** to the Indian Institute of Technology Delhi, for the award of the Degree of **Doctor of Philosophy**, is a record of the bona fide research work carried out by him under my supervision and guidance. The thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted either in part or in full to any other University or Institute for the award of any degree or diploma to the best of my knowledge.

Prof. Shaunak Sen
Department of Electrical Engineering,
Indian Institute of Technology Delhi.
(Supervisor)

ACKNOWLEDGEMENTS

I would like to begin by thanking my thesis supervisor, Prof. Shaunak Sen for all his guidance, advice, support, encouragement, patience, and enthusiasm over the years during the period of my Ph.D. study here at IIT Delhi. I would also like to give special thanks to my SRC members: Prof. S. Janardhanan, Prof. Deepak U. Patil, and Prof. Hariprasad Kodamana for their support and valuable comments. I would always be grateful to Prof. Priyank Srivastava, Prof. Subashish Datta, Prof. Shubhendu Bhasin, and Prof. I. N. Kar for their help and support.

A Ph.D. is not all about writing a thesis; rather it involves learning new things and, polishing and strengthening the existing knowledge base. Nevertheless, this journey would not have been pleasant without the warm reception of all my lab mates and staff members of the Control and Automation group of IIT Delhi. On a personal note, I would like to thank my family for their constant support and encouragement, especially in the difficult times.

Finally, I thank the almighty for the blessings which have made this thesis possible.

Krishan Kumar Gola

ABSTRACT

In the developing area of synthetic biology, researchers engineer biomolecular circuits within living cells to manipulate and control cellular behaviour for a variety of applications, such as energy production, environmental management, and medical applications. Feedback control strategies, rely on sensors, and efficient estimation strategies are employed to achieve the desired control in biomolecular systems.

RNA thermometers, which are temperature sensors, generate temperature responses with specific thresholds and sensitivities. Researchers have extensively studied the structure and sequences of RNA thermometers, which are mRNAs whose translation rate varies with temperature. The gradual melting of the ribosome-binding site (RBS) at increased temperatures is a key part of this thermosensing process, but the parameters on which the melting of RBS depends is generally unclear. Designing an RNA thermometer that exhibits a precise response, including a specific peak sensitivity and threshold temperature, poses a considerable challenge. We investigated the tradeoffs present in the input-output response of RNA thermometers using mathematical models such as the two-state model and its variants. We found that peak sensitivity and threshold temperature show tradeoffs. To further validate the consistency of the above-mentioned tradeoffs in more realistic models, we used NUPACK, a web-based tool designed for the thermodynamic analysis of nucleic acids. By calculating the melt profiles of synthetic RNA thermometers through this thermodynamics-based server, we found that the results were consistent with the stated constraints. To check for the above trends in the experimental measurements of RNA thermometers, we designed two libraries on either side of the ribosome binding site and experimented with different temperatures. We found that thermometer activities change with temperature and are different for all thermometers. We also found that peak sensitivity and threshold temperature persist in experimental measurements as well.

RNA thermometers have been used as sensors in multiple contexts for gene regulation. However, they can be subject to noise, inaccuracies, or limitations. Like Kalman filters or observers, estimation techniques are required to estimate the true system states.

Estimation plays a crucial role in understanding and analyzing biomolecular systems, as these systems are often complex and difficult to study directly due to their dynamic and intricate interaction. Stochastic differential equation-based models have been used in biomolecular contexts that can have a state-dependent process noise covariance. The choice of the process noise covariance is an important parameter in the design of a Kalman Filter for state esti-

mation, and the theoretical guarantees of updating the process noise covariance as the state estimate changes are unclear. We investigated this issue using the best linear unbiased estimator (BLUE), and Newton’s method by minimizing a weighted least squares cost function for the systems where process noise depends on the state. We found that the BLUE-based algorithm was an optimal linear filter for linear system dynamics when the process noise depends on the linear/affine function of state in a square-root-like fashion. We generalized these results by minimizing a positive definite quadratic function using Newton’s method, which removes the restriction on process noise dependent on the state being a linear/affine function of the state. We generalized these results to nonlinear systems with arbitrary process noise dependence, showing that this algorithm minimizes a quadratic approximation to a least-squares cost weighted by the noise covariance. We considered one example to illustrate these results. This study could aid in the manipulation of biomolecular systems, contribute to the development of ideas, and pave the way for future research in synthetic biology and related fields.

Keywords: Biomolecular Systems, RNA Thermometers, Sensitivity, Threshold Temperature, Best Linear Unbiased Estimator, Newton’s method, Chemical Langevin Equation (CLE), Mean-Square Error (MSE).

सार

कृत्रिम जीवविज्ञान के उभरते हुए क्षेत्र में वैज्ञानिक जीवित कोशिकाओं के भीतर जैव-अणु परिपथ का निर्माण करते हैं, जिससे कोशिकाओं के व्यवहार को नियंत्रित और निर्देशित किया जा सके, इसका उद्देश्य ऊर्जा उत्पादन, पर्यावरणीय संरक्षण, और चिकित्सा जैसे विविध क्षेत्रों में उपयोगी समाधान विकसित करना है। इन जैव प्रणालियों में वांछित नियंत्रण प्राप्त करने के लिए फीडबैक नियंत्रण रणनीतियाँ, संवेदक, तथा सटीक और प्रभावी अनुमान तकनीकों का सहारा लिया जाता है।

आरएनए थर्मोमीटर, जो तापमान संवेदी अणु होते हैं, विशिष्ट सीमा और संवेदनशीलता के साथ तापमान के प्रति प्रतिक्रियाएँ उत्पन्न करते हैं। वैज्ञानिकों ने आरएनए थर्मोमीटरों की संरचना और अनुक्रमों का व्यापक रूप से अध्ययन किया है, ये थर्मोमीटर वास्तव में ऐसे एमआरएनए होते हैं जिनकी अनुवाद दर तापमान के अनुसार परिवर्तित होती है। तापमान बढ़ने पर राइबोसोम-बाइंडिंग साइट (आरबीएस) का क्रमिक रूप से पिघलना इस ताप संवेदी प्रक्रिया का एक प्रमुख घटक है, किन्तु आरबीएस के पिघलने की प्रक्रिया किन-किन विशिष्ट कारकों पर निर्भर करती है, यह अभी भी स्पष्ट रूप से ज्ञात नहीं है। एक ऐसा आरएनए थर्मोमीटर डिजाइन करना, जो किसी विशेष सीमा तापमान और अधिकतम संवेदनशीलता के साथ सटीक प्रतिक्रिया प्रदान करे, एक जटिल चुनौती प्रस्तुत करता है। इस समस्या की प्रकृति को समझने के लिए, हमने आरएनए थर्मोमीटरों की इनपुट-आउटपुट प्रतिक्रियाओं में मौजूद ट्रेडऑफ़्स का अध्ययन गणितीय मॉडलों जैसे कि दो-स्थिति मॉडल और उसके विविध रूपों के माध्यम से किया। हमारे अध्ययन से यह ज्ञात हुआ कि अधिकतम संवेदनशीलता और सीमा तापमान के बीच स्पष्ट द्वंद्व विद्यमान हैं। इन सिद्धांतों की यथार्थवादी मॉडल्स में संगतता की पुष्टि हेतु हमने न्यूपैक नामक एक वेब-आधारित प्लेटफ़ॉर्म का उपयोग किया, जो न्यूक्लिक एसिड के थर्मोडायनामिक विश्लेषण के लिए विकसित किया गया है। इस प्लेटफ़ॉर्म की सहायता से जब हमने कृत्रिम आरएनए थर्मोमीटरों की मेल्ट प्रोफाइल का विश्लेषण किया, तो पाया कि थर्मोडायनामिक मॉडल आधारित निष्कर्ष भी उपरोक्त ट्रेडऑफ़्स के अनुरूप हैं। इसके बाद, हमने इन प्रवृत्तियों की पुष्टि के लिए प्रयोगात्मक परीक्षण किए। हमने राइबोसोम-बाइंडिंग साइट के दोनों ओर दो आरएनए लाइब्रेरीज़ डिजाइन कीं और विभिन्न तापमानों पर उनके व्यवहार का परीक्षण किया। हमें यह पाया कि प्रत्येक थर्मोमीटर की गतिविधि तापमान के अनुसार भिन्न होती है, और हर आरएनए थर्मोमीटर की प्रतिक्रिया अद्वितीय होती है। इसके अतिरिक्त, यह भी पुष्टि हुई कि अधिकतम संवेदनशीलता और सीमा तापमान जैसी द्वंद्व प्रवृत्तियाँ प्रयोगात्मक परिणामों में भी विद्यमान रहती हैं।

आरएनए थर्मोमीटर का उपयोग जीन नियमन के लिए विभिन्न संदर्भों में संवेदक के रूप में किया गया है। तथापि, ये शोर, त्रुटियों या सीमाओं के प्रति संवेदनशील हो सकते हैं। कालमैन फिल्टर्स या प्रेक्षकों की तरह, वास्तविक प्रणाली अवस्थाओं का अनुमान लगाने हेतु अनुमान तकनीकों की आवश्यकता होती है।

अनुमान जैव-अण्विक प्रणालियों को समझने और विश्लेषण करने में एक महत्वपूर्ण भूमिका निभाता है, क्योंकि ये प्रणालियाँ अक्सर जटिल होती हैं और उनकी गतिशील और जटिल पारस्परिक क्रियाओं के कारण सीधे अध्ययन करना कठिन होता है। जैव-अण्विक संदर्भों में स्टोकास्टिक डिफरेंशियल समीकरण-आधारित मॉडलों का उपयोग किया गया है, जिनमें स्थिति-निर्भर प्रक्रिया शोर सहप्रसरण हो सकती है। प्रक्रिया शोर सहप्रसरण का चयन कालमैन फ़िल्टर डिज़ाइन में एक महत्वपूर्ण पैरामीटर होता है, और जैसे-जैसे स्थिति का अनुमान बदलता है, प्रक्रिया शोर सहप्रसरण को अद्यतन करने की सैद्धांतिक गारंटी स्पष्ट नहीं होती है। हमने इस समस्या की जाँच सर्वश्रेष्ठ रैखिक निष्पक्ष अनुमानक (बीएल्यूई) और न्यूटन विधि का उपयोग करके की, जिसमें हमने उन प्रणालियों के लिए वेटेड लीस्ट स्क्वेयर कॉस्ट फ़ंक्शन को न्यूनतम किया जहाँ प्रक्रिया शोर स्थिति पर निर्भर होता है। हमने

पाया कि बीएलयूई-आधारित एल्गोरिद्म रैखिक प्रणाली गतिकी के लिए एक इष्टतम रैखिक फिल्टर है, जब प्रक्रिया शोर स्थिति के रैखिक/एफ़ाइन फ़ंक्शन पर वर्गमूल जैसे रूप में निर्भर करता है। हमने इन परिणामों को न्यूटन विधि का उपयोग करके एक धनात्मक परिभाषित द्विघात फ़ंक्शन को न्यूनतम करके सामान्यीकृत किया, जिससे प्रक्रिया शोर की स्थिति पर निर्भरता को रैखिक/एफ़ाइन होने की बाध्यता समाप्त हो जाती है। हमने इन परिणामों को गैर-रैखिक प्रणालियों तक विस्तारित किया, जहाँ प्रक्रिया शोर की निर्भरता मनमानी हो सकती है, और यह दर्शाया कि यह एल्गोरिद्म एक द्विघात अनुमानित लीस्ट-स्केयर कॉस्ट को न्यूनतम करता है जिसे शोर सहप्रसरण द्वारा वेट किया गया होता है। हमने इन परिणामों को स्पष्ट करने के लिए एक उदाहरण पर विचार किया। यह अध्ययन जैव-अण्विक प्रणालियों के नियंत्रण में सहायक हो सकता है, विचारों के विकास में योगदान दे सकता है, और कृत्रिम जीवविज्ञान तथा संबंधित क्षेत्रों में भविष्य के शोध के लिए मार्ग प्रशस्त कर सकता है।

मुख्य शब्द: जैव-अण्विक प्रणालियाँ, आरएनए थर्मामीटर, संवेदनशीलता, सीमा तापमान / थ्रेशोल्ड, सर्वश्रेष्ठ रैखिक निष्पक्ष अनुमानक, न्यूटन विधि, रासायनिक लांगेविन समीकरण, औसत-वर्ग त्रुटि ।

Contents

Certificate	i
Acknowledgements	iii
Abstract	v
List of Figures	xiii
List of Tables	xvii
List of Abbreviations	xix
List of Symbols	xxi
1 Introduction	1
1.1 Background	1
1.1.1 Literature Review	1
1.1.2 Motivation	6
1.2 Problem Statement	8
1.3 Organization of the Thesis	9
2 Tradeoffs	11
2.1 Introduction	11
2.2 Mathematical Models	11
2.2.1 Two-State Model	11
2.2.2 Sequential n -State Model	17
2.2.3 Branched Model	19
2.3 NUPACK Data	21
2.3.1 RBS Melt Profile	22

2.3.2	RBS Sensitivity Profile	22
2.3.3	Peak Sensitivity and Threshold Temperature	23
2.3.4	Melt Profile Fits	23
2.4	Summary	28
3	Experimental Evidence	29
3.1	Introduction	29
3.2	Design	29
3.2.1	Procedure	29
3.2.2	Hypothesis	30
3.2.3	Analysis	32
3.3	Materials	34
3.4	Methods	34
3.4.1	Libraries	34
3.4.2	Experimental Procedure	35
3.5	Results	36
3.6	Practical Guidelines for Designing RNA Thermometers	41
3.7	Conclusion	43
4	Estimation: Newton Method Approach to State-Dependent Process Noise	
	Framework	45
4.1	Introduction	45
4.2	Preliminaries	48
4.2.1	Minimum Mean-Square Estimation	48
4.2.2	MMSE for Gaussian Vectors	49
4.2.3	Best Linear Unbiased Estimator	52
4.2.4	Newton's Method	54
4.2.5	Weighted Least Squares	54
4.3	Estimation	56
4.3.1	Linear Discrete-Time System	56
4.3.2	Stochastic Non-linear System	65
4.3.3	Examples (scalar case)	72
4.3.4	Simulation Results	73
4.3.5	Performance of Filter	74

4.4	Example with Experimental Data	76
4.5	Conclusion	77
5	Conclusions and Future Directions	79
5.1	Summary of Contributions	79
5.2	Future Directions	80
5.3	Relevance of Work	80
	Bibliography	83
	Appendix	93
	List of Publications	119
	Biodata	121

List of Figures

1.1	Estimation and sensors are the two key parts of biomolecular system design . . .	2
1.2	Working principle of RNA thermometer	2
2.1	Schematic representation of two states — folded and unfolded — of an RNA thermometer.	12
2.2	Graphical representation shows there is one maximum point	15
2.3	Two-state model: varying parameter b	16
2.4	Schematic representation of n -states — folded and unfolded — of an RNA thermometer.	17
2.5	Sequential model: varying parameter b	18
2.6	Schematic representation of branched - folded and unfolded - of an RNA thermometer	19
2.7	Branched model: varying parameter b	20
2.8	Black line with circles, blue line with squares, red line with diamonds represents the peak sensitivity for the two-state model, the sequential variant ($n = 3$), and the branched variant, respectively	21
2.9	Melt profile and RBS melt profile for U2 thermometers and its mutations.	22
2.10	Blue squares, red circles, and black diamonds represent the thermometers with one peak, two peaks, and three peaks, respectively.	23
2.11	LSE fit plots: Black lines are for RBS response and sensitivity profile, and red lines are for LSE fit. Green and blue lines for two peak-fitted parameters used in two-state models only	25

2.12	Tradeoffs in computed melt profiles.: Circles, squares, and diamonds represent the peak sensitivity and threshold of fits to the melt profiles having one peak, two peaks, and three peaks respectively. The fill colour is the normalized value of the parameter b computed from the expression $(b - \min_b)/(\max_b - \min_b)$, where $\max_b$ and $\min_b$ are the maximum and minimum over the entire set of b values, respectively. a). Thin black lines represent a grid of lines generated from Equation 2.3 for different values of b . b) peak sensitivity and threshold coming from the NUPACK-generated melt profile rather than the fit.	26
2.13	Tradeoffs in computed melt profiles for U0 - U12 thermometers. (a) melt profile (b) RBS melt profiles (c) RBS sensitivity profiles (d) Blue squares, red circles, and black diamonds represent the thermometers with one peak, two peaks, and three peaks, respectively (e) Peak sensitivity vs threshold temperature for fit data (f) Peak sensitivity vs threshold temperature for NUPACK data	27
3.1	NUPACK analysis for U2 thermometer - as the temperature increases Minimum Free Energy Structure and probabilities of RBS base changes (a) MFE structure at 29°C (b) MFE structure at 37°C (c) probabilities of RBS at 29°C (d) probabilities of RBS at 37°C.	31
3.2	NUPACK analysis - as the one base pair change leads to change in Minimum Free Energy Structure, free energy of secondary structure, and probabilities of RBS bases (a) MFE structure 22O at 29°C (b) MFE structure of 34O at 29°C (c) probabilities of RBS of 22O at 29°C (d) probabilities of RBS of 34O at 29°C.	33
3.3	Schematic of a plasmid containing an RNA thermometer and other essential elements	34
3.4	Experimental procedure	36
3.5	NUPACK data analysis for designed libraries: (a) melt profile (b) RBS melt profiles (c) RBS sensitivity profiles (d) Peak sensitivity vs threshold temperature – Black and red dots represent the library 1 and library 2, respectively (e) Peak sensitivity vs threshold temperature for fit data (f) Peak sensitivity vs threshold temperature for NUPACK data. Thin black lines represent a grid of lines generated from Equation 2.3 for different values of b	37
3.6	F/OD of library 1 and library 2 thermometers experiment performed on three different days	39

3.7	F/OD of library 1 and library 2 thermometers by taking the average of all three day	40
3.8	Sample LSE fits. The black line is the RBS melt profile, and the red line is the RBS melt profile fits (a) Library 1 (b) Library 2	40
3.9	Tradeoffs in experimental measurements (a) Black and red dots represent the peak sensitivity, and black and red bars represent the associated error bar, which represents the variability across experimental replicates for library 1 and library 2 thermometers (b) The peak sensitivities and threshold temperature are normalized with the corresponding parameters. Thin black lines represent a grid of lines generated (Equation 2.3) for different values of b	41
3.10	The flowchart of the practical guideline for designing RNATs that take into account the above performance tradeoffs	42
4.1	Time trajectories for one thermometer	46
4.2	Protein production-degradation system	46
4.3	State estimate for 100 iteration for a) $\beta = 0.1$, b) $\beta = 1$, c) $\beta = 10$ and d) state-dependent noise framework	74
4.4	Error mean and standard deviation for a) $\beta = 0.1$, b) $\beta = 1$, c) $\beta = 10$ and d) state dependent noise framework	75
4.5	Mean square error for a) $\beta = 0.1$, b) $\beta = 1$, c) $\beta = 10$ and d) state-dependent noise framework	76
4.6	Dots represent the experimental data, the dashed line represents the estimated state from the Kalman filter with $\beta = 0.1$, and the solid line represents the estimated state from a filter using a state-dependent process noise covariance update framework.	77

List of Tables

1.1	RNA thermometer list and their threshold temperature	7
3.1	U0-U12 thermometers with their corresponding ‘ b ’ values	32
4.1	Quantitative comparison between state-dependent process noise covariance update framework and Kalman filter with different fixed process noise covariance. .	76

List of Abbreviations

RBS	Ribosome Binding Site
RNATs	RNA Thermometers
5'-UTR	5'- Untranslated Region
ROSE	Repression of Heat Shock Gene Expression
<i>E. coli</i>	<i>Escherichia coli</i>
ODE	Ordinary Differential Equation
SDE	Stochastic differential Equation
CME	Chemical Master Equation
CLE	Chemical Langevin Equation
EKF	Extended Kalman Filter
MMSE	Minimum Mean-Square Estimation
WLS	Weighted Least Square
HEKF	Hybrid Extended Kalman Filter
NLS	Nonlinear System

List of Symbols

g_{max}	Maximum sensitivity
R_f	Concentration of RNA in folded state
R_u	Concentration of RNA in unfolded state
R_T	Concentration of total RNA present at any instant of time
k_1 & k_2	Reaction rate constant
T	Absolute Temperature in degree Kelvin
E_1 & E_2	Activation energy in Joule
k	Boltzmann constant
y	Response/fraction of basis unpaired
a & $b > 0$	Enthalpic and entropic constants
G	Sensitivity
G_{max}	Peak sensitivity
$T_{Threshold}$	Threshold temperature in Kelvin
x	vector x
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^n$	Real Space of dimension n
$\mathbb{R}^m$	Real Space of dimension m
$\mathbb{R}^{n \times n}$	Real Matrix of dimension $n \times n$
u	Control input
y	Output vector
v_k	Process noise
w_k	Measurement noise
Q_k	Process noise covariance
R_k	Measurement noise covariance
$\hat{x}_{k k-1}$	Estimate of x_k based on measurements upto $y_1, y_2, \dots, y_{k-1}$
$\hat{x}_{k k}$	Estimate of x_k based on measurements upto $y_1, y_2, \dots, y_k$
$P_{k k-1}$	Covariance of the estimation error of $\hat{x}_{k k-1}$
$P_{k k}$	Covariance of the estimation error of $\hat{x}_{k k}$

