UNIVERSITY OF CALIFORNIA SAN DIEGO

System Identification and Modeling for Control of a High Power Drive Laser

A thesis submitted in partial satisfaction of the
requirements for the degree Master of Science

in

Engineering Sciences - Mechanical Engineering

by

Soren Eric Almquist

Committee in charge:

Professor Robert Bitmead, Chair
Professor Raymond A. de Callafon
Professor Boris Kramer

2023

The Thesis of Soren Eric Almquist is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I'd like to give my sincere thanks and show of appreciation to everyone who has helped me succeed in my graduate studies endeavor. . .

Eric and Mary Almquist, who gave me all the tools I need for success,

Kayla Howie, who pushed me to believe in myself no matter the challenge,

Leo Garnar-Wortzel and Khalyd Clay, who kept the times fun over this journey,

Professor Robert Bitmead, my advisor and committee chair, who guided me to success,

My professors Raymond de Callafon, Mauricio Oliveira, Boris Kramer, and Tom Bewley,

Holger Schmidt, from the Applied Optics lab at UC Santa Cruz,

Andrei Litvak and Carmen Zoldesi, my leadership at ASML,

Daniel Riggs and Wayne Dunstan, who inspired me to take on this endeavor,

The entire Source Performance team at ASML, who helped me maintain balance over these years

To everyone, thank you!

ABSTRACT OF THE THESIS

System Identification and Modeling for Control of a High Power Drive Laser

by

Soren Eric Almquist

Master of Science in Engineering Sciences - Mechanical Engineering

University of California San Diego, 2023

Professor Robert Bitmead, Chair

The EUV light source is a key enabler of modern photolithographic computer chip manufacture. In the light source, a high-power drive laser irradiates a tin fuel source to generate an EUV light-producing plasma. Regulation of the drive laser power is critical to ensure that the EUV light source generates the precise amount of light needed for the photolightographic process. The laser power can be adjusted by modulating the excitation medium, however, the amplification process is highly dynamic and exhibits transient behavior on short timescales. These fast timescale dynamics are important to understand as the EUV light source power must be tightly regulated. In this work, these dynamics are investigated towards the objective of developing a control system framework to regulate the drive laser power. An experiment to

characterize the drive laser was carried out on a prototype drive laser system, and the results used to derive a model of the drive laser amplification process using system identification. A filtered output error model is developed which results in accurate simulation of the drive laser dynamic behavior. This model is then demonstrated to yield good performance for feed-forward control, using a model inversion framework. These results show that the modeling process is successful in generating a control-relevant model of the high power drive laser system.

Chapter 1

The EUV light source and high-power drive laser system

1.1 Overview of the thesis

Nano-scale electronic circuit photolithography is the key process that enables the ubiquity and utility of modern technology. As computer chips are shrunken to to ever-smaller dimension, technologies become democratized and distributed - computations that historically required mainframe computers or expensive hardware can now be performed on an everyday mobile phone. The performance capability of a computer chip is inverse to the dimension of the electronic circuits that constitute it; in order to squeeze these circuits to-ever smaller dimension, the photolithographic process must etch smaller and smaller features, which requires more and more rigorous control over the stability of the process. Higher-fidelity control over each component process within modern computer chip manufacturing enables wider distribution of the benefits of technological advancement and accelerates the rate at which technology impacts our lives.

In this thesis, I will describe a key dynamic process in modern photolithography and investigate various contributions to the instability of the process. A system-identification framework is used to develop a control-relevant model of the process, and control designs are explored that can be used to improve the process's stability. This work thereby indicates a direction towards improving the capability of the overall photolithographic process to create ever-smaller

computer chip circuits.

1.2 Photolithography

In modern computer chip manufacture, the photolithographic process describes using light to etch circuit patterns onto a silicon substrate; these circuit patterns are then stacked and interconnected to form a computer chip. The process encompasses generating a high-power beam of light of a given wavelength, and directing the light so that it is incident onto a photosensitive chemical polymer ("resist") applied to the silicon wafer. In between the light source and the substrate is a mask ("reticle") with the desired printing pattern etched in it; the light photons passing through the mask then only contact the regions of the resist corresponding to the mask's printing pattern. After development, the pattern of resist that remains on the substrate is a scaled copy of the mask's original pattern. This highly selective control over the exposure of light onto the resist then allows spatio-selective ion implantation, metal deposition, or material etching, all of which are used to generate extremely precise nano-scale structures in the silicon wafer substrate. These structures are then patterned and stacked to a dizzying density; modern computer chips pack around 10^8 transistors per mm^2. With higher transistor density, and therefore smaller circuit size, the performance of the computer chip improves in terms of compute speed per energy consumption.

The minimum dimension of the exposure pattern that can be etched into the resist is determined by the wavelength of light used, as described by the Rayleigh equation

$$CD = \frac{k\lambda}{NA}$$

where CD is the "Critical Dimension" of the exposed feature, k is a coefficient that includes properties of the resist and manufacturing process, λ is the wavelength of the incident light, and NA is the Numerical Aperture of the optical system. The most advanced modern photolithography processes use a 13.5nm "Extreme Ultra-Violet" (EUV) light generated by a tin plasma. This

light is directed into the "scanner," which focuses the light with a given numerical aperture onto the silicon wafer substrate.

The wavelength of light determines the feature size of the printed pattern. However, as important, is the power of the light source. The power is proportional to the number of photons, of a given wavelength, generated per time. To manufacture a feature with a given dimension, the lithographic manufacturing process adjusts the photons per area incident on the resist; this metric is the applied "dose." A higher power light source can achieve the target dose in less time, which improves the throughput of the process. Equally important is the stability of the applied dose, and therefore the stability of the light source power output. A high degree of stability is necessary to ensure that the desired pattern is printed with high fidelity. Optimizing both the power and stability of the light source requires precise control over the mechanisms and components that generate the EUV light.

ASML is the leading photolithography equipment producing company in the world, and the only company that has commercialized EUV light sources and scanners. The work described in this thesis is performed on equipment and data that is wholly owned by ASML, in a project that was a collaboration between UC San Diego and ASML. All data described in this thesis has been scaled, and the units removed, so as to protect ASML intellectual property.

1.3 EUV light source

In order to generate 13.5nm EUV light with sufficient power to meet the throughput requirements, the EUV light source uses a Laser-Produced Plasma (LPP) design. Generating a single pulse of EUV photons is achieved by irradiating a $30\mu m$ tin fuel source (a "droplet") with a 30kW infrared (IR) laser. The tin droplet is shot across the vessel at high speed by the droplet generator and is intercepted precisely by the infrared laser pulses, with micron spatial precision and microsecond timing precision. The laser pulses then ionize the droplet into a tin plasma with highly-energized free electrons. As the electrons recombine with the tin ions, they produce

13.5nm EUV photons. These photons are then captured by a multi-layer collector mirror, which reflects the light into the scanner. The collector mirror is parabolic, with the plasma placed at the mirror primary focus (PF), and the input to the scanner placed at the intermediate focus. With this positioning, the light generated at the primary focus point, and directed towards the collector mirror, is then reflected to the intermediate focus point and into the scanner, to then be focused onto the silicon wafer substrate.

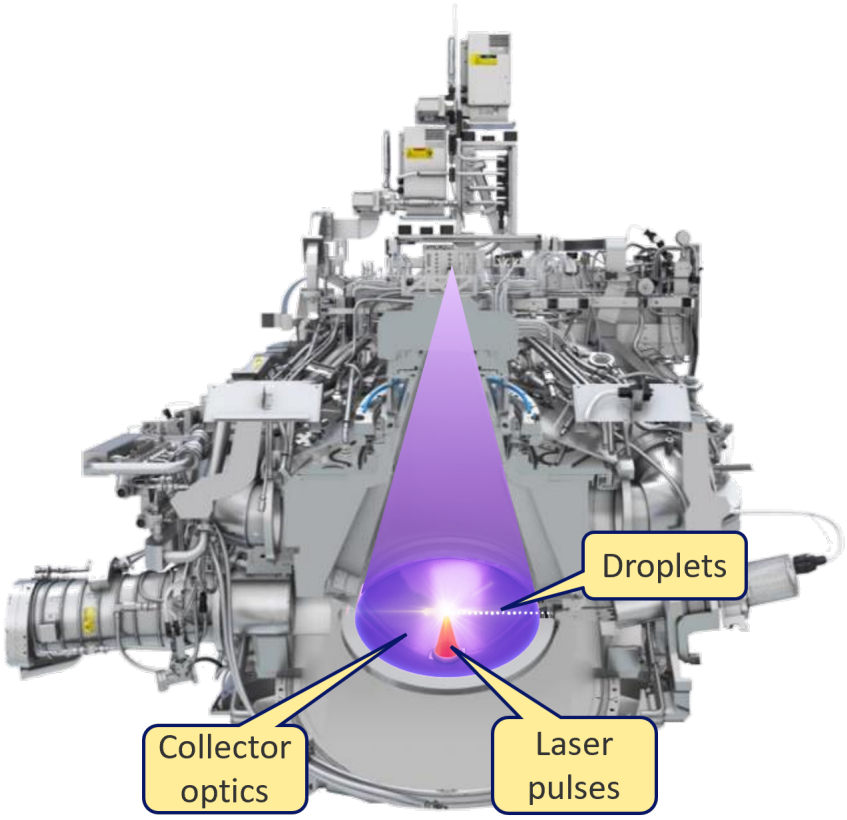


Figure 1.1. EUV light source vessel. Droplets, in white, travel from right to left, and are intercepting by the infrared laser pulses, in red, at the primary focus of the collector optics. The EUV light generated, in purple, is focused by the collector optics to the intermediate focus, towards the scanner, which is not shown. [1].

The EUV-emitting plasma generation process is repeated at least 50,000 times per second. This requires that both individual tin droplets and individual laser pulses are generated at most every 20 μs, and synchronized such that they arrive at the focal point of the collector mirror with

precision timing. Note that the 50 kHz rate is specified as the standard rate for commercial light sources, however in a laboratory setting the rate can be varied. The laser and droplet interception process is highly dynamic due to the intrinsic variability of the inputs. Variability in the spacing between droplets, the relative positioning between the droplets and the laser pulses at primary focus (PF), and the energies of the laser pulses will all impact the stability of the EUV generation process. Each of these constituent inputs must be tightly controlled to maintain high overall EUV power with low variability.

Within a 20 μs cycle of the EUV generation process, the infrared laser system generates a pair of laser pulses, the Pre-Pulse (PP) and the Main-Pulse (MP). The Pre-Pulse intercepts the droplet first, causing the spherical droplet to turn into a disc shape and decrease in density. Microseconds later, the Main-Pulse intersects the expanding disc and ionizes it, generating the EUV light-producing plasma. The Pre-Pulse is instrumental to preparing the morphology of the tin droplet for efficient coupling with the Main-Pulse. The two pulses are generated by separate low-power seed lasers, and are then amplified together in the amplification chain. This is described in further detail in the next section.

In this thesis, the stability of the amplified laser pulse energies is investigated. Sources of instability to the laser energies are studied and a dynamical model of the laser pulse energies is developed, focusing on the model's utility for control of the laser system. A control system framework is suggested that should improve the pulse-to-pulse stability of the lasers.

1.4 High-power drive laser

The generation of high-energy infrared laser pulses is achieved via the "High-power drive laser" system, or drive laser (DL) for short. The laser system is of a "MOPA (Master Oscillator Power Amplifier) + Pre-pulse" architecture, which generates, amplifies, and transports the laser beams into the EUV vessel, where they are incident onto the tin droplets. The laser generation process occurs as follows, as shown in the subsequent figure:

Pulses of infrared laser light are generated by the master oscillators in the seed system (at left) at 50,000 times per second. The pulses are brought into the power amplification stage, where they are amplified via an energized medium in a series of four Power Amplifiers. When the input seed pulses of IR light enter the amplifiers, they stimulate the energized medium to lase and generate more IR photons, diminishing the energy stored in the medium. This amplification process continues as the pulses propagate through the power amplifiers, stimulating more lasing and increasing the power of the IR laser dramatically. At the output of the power amplification stage, the laser reaches up to 30kW of steady-state power. The beams are then brought to the EUV vessel where they are focused onto the tin droplets. The oncoming droplets are synchronized to the laser pulses, such that the droplets are precisely intercepted by the laser pulses.

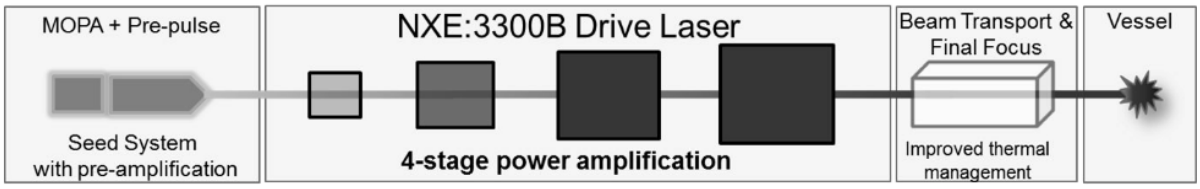


Figure 1.2. Schematic of the high-power drive laser system, showing the beam paths of the Main-Pulse, Pre-Pulse, and Reverse (MP+PP) laser light beams [4].

1.5 Laser pulse amplification

The amplification of the infrared laser light occurs due to the balance between the following three processes:

1.5.1 Pumping

Within the power amplifiers, there is an excited gain medium through which the laser pulses pass. The excited medium can be energized by promoting the medium's electrons from their ground state, a low energy orbit, into a higher-energy state. This process is called "pumping." As the pumping rate increases, the number of electrons inhabiting the higher energy level increases. If the number of electrons in the higher level is higher than that in the lower level, the gas medium is said to have a "population inversion," which is necessary to achieve lasing.

1.5.2 Stimulated emission

The power amplification of the laser light is achieved via the fundamental "laser" principle, **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation. When the input infrared "seed" laser pulse enters the power amplifier, it encounters the energized medium, and stimulates the energized electrons to collapse to their ground level, emitting more IR photons as a result. This process is "stimulated emission."

The energy of the emitted photons (and therefore the wavelength) is determined by the bandgap between the ground state and the excited energy levels of the electrons. These emitted IR photons are of the same wavelength and coherent with the input seed laser IR photons (i.e. they are in-phase and propagate in the same direction as the input pulses); in this way, the input laser light pulse power has been amplified. The amplification is proportional, to a degree, to the power of the input laser pulse - the more input infrared photons propagate through the energized gas, the more likely it is that a given excited gas molecule will undergo stimulated emission. However, the amplification can only be proportional up to the point where the entire population inversion is depleted, at which point no further amplification can occur. In total the output power of the amplified light is proportional to the gain of the lasing medium (the population inversion), determined by the pumping, in addition to the power of the input seed laser pulse.

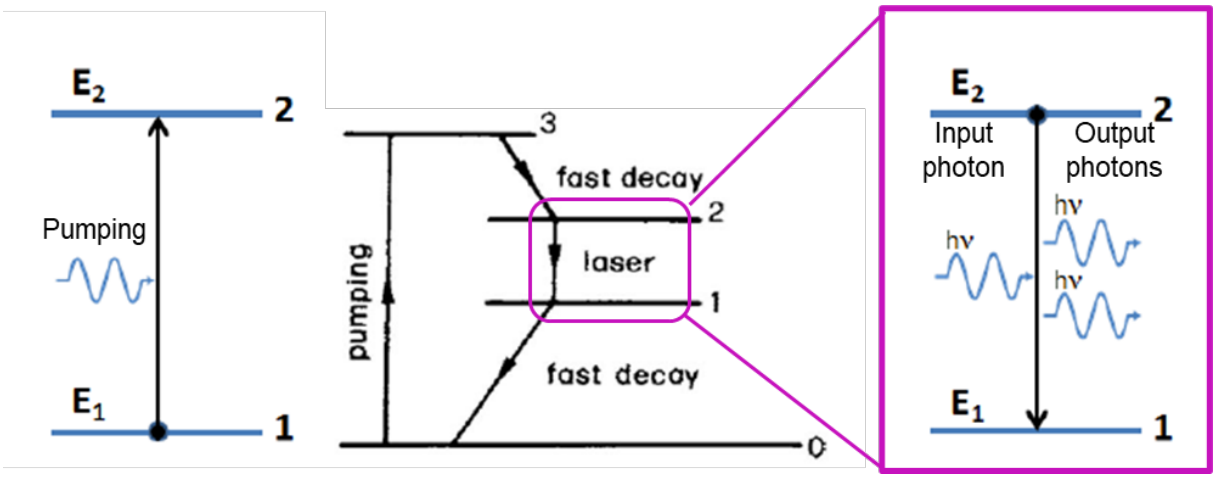


Figure 1.3. Pumping and stimulated emission mechanisms in a multi-energy level electron system. The pumping mechanism excites electrons from the ground E1 state to the excited E2 state, in the left figure. An input photon, $h\nu$, stimulates the excited electron to collapse from the excited state E2 to the lower-energy state E1, emitting a second photon with energy $h\nu$ in the process, shown in the right figure

1.6 Drive laser amplification model

The amplification of the seed laser pulses is the key process that allows for high-power EUV light generation. The power of the EUV light is proportional to the power of the IR laser. Hence, the IR laser light power is amplified as high as possible while maintaining low variability. This thesis will focus on the amplification dynamics of the laser light through the serial amplifiers, specifically, how the output laser energy and the gain dynamics of the amplifiers are affected by lasing medium excitation. The excitation of the medium can be controlled via an input signal dubbed the Gain Command. The amplified laser pulse energy is measured for each laser pulse exiting the amplification stage.

The focus of this thesis is to develop a model of of the laser pulse amplification process in an EUV source drive laser. The dynamic model that will be studied is the relationship between this input Gain Command and the output amplified laser energy.

Chapter 2

Experimental data

2.1 Experimental method

The dynamic response of the laser amplification process was tested on a drive laser connected to a prototype EUV light source located at ASML San Diego. Note that the data presented in this section, where applicable, have been re-scaled to anonymize the physical units of each measurement, to protect intellectual-property. The re-scaling does not affect the validity of the identification or modeling results. The scaling factors have been preserved for future in-house reference.

The input tested was a square wave input to Gain Command. Data from the controller input signal was collected at a rate of one data point per shot, and measurements from the amplified laser output were recorded and synchronized per shot.

2.2 Time-domain analysis of dynamic behavior

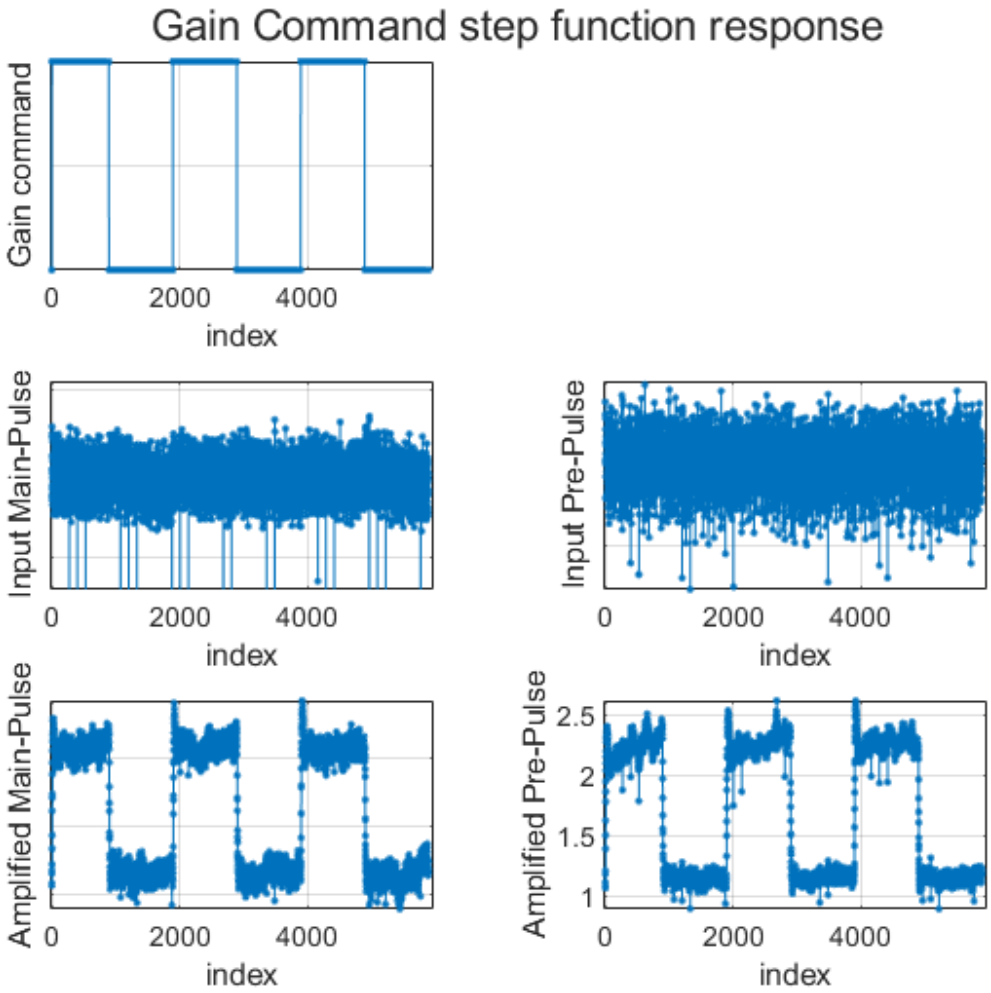


Figure 2.1. The drive laser amplification response to adjusting the power of the Gain Command. This data is from one experiment where Gain Command was stepped several times in succession. The amplified laser energies increase as a result of increasing the gain command. The input seed laser energies remain constant

The above Figure 2.1 shows three example cycles of the Gain Command step sequence. The input steps yield steps in the amplified energies of the Main-Pulse and Pre-Pulse commensurately. The inputs to the amplification chain are unaffected by the Gain Command input.

2.2.1 Second-order response

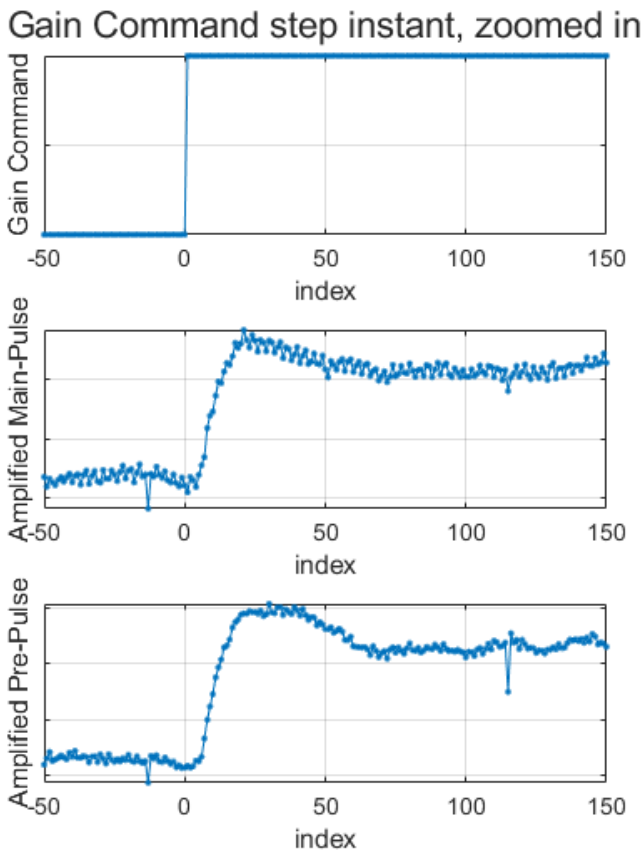


Figure 2.2. Example trace of transient response of amplified laser pulse energies in response to Gain Command. X-axis is in shot-number, with Gain Command step at index 0

Zooming into an example trace (Figure 2.2) shows the transient response of the amplified pulse energies to the Gain Command step. This response is roughly "second-order" as the measured energy overshoots and then settles to the steady-state level. From the moment of the commanded step change in the Gain Command (at index 0), the amplified laser pulse energies begin to increase with a 5-index delay. There is a gradual transient as the amplified pulse energies increase up to reach an overshoot , then the energies then rebound and stabilize, reaching steady-state at 70 indices after the commanded step. The entire transient response occurs within the millisecond time range after the commanded step impulse. This timescale is relevant for control

of the IR laser pulse energy, as millisecond-timescale errors in the EUV energy produced by the light source will affect the exposed feature size on the resist.

2.2.2 Ensemble performance and variability

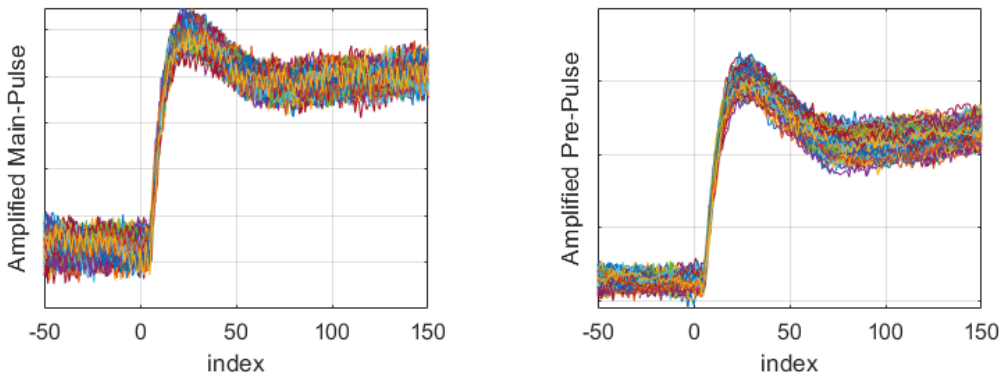


Figure 2.3. The ensemble of drive laser amplification responses to adjusting the Gain Command. Each trace is a single response to the Gain Command step-function input, from 50 indices prior to 150 indices after the commanded step. 164 total step responds are overlaid

The ensemble response is shown by overlaying the 164 traces of the periodic step-function input. Across the ensemble, the second-order transient response is found to be very consistent between the traces. Across the ensemble however the traces display variability in the steady-state levels on either side of the Gain Command step input, trace-to-trace. The amplified Main-Pulse and Pre-Pulse energies show variability, respectively, in the steady-state across the traces, with constant Gain Command input. This variation trace-to-trace is deleterious to the stability of the EUV power delivered to the resist.

2.2.3 High-frequency variability

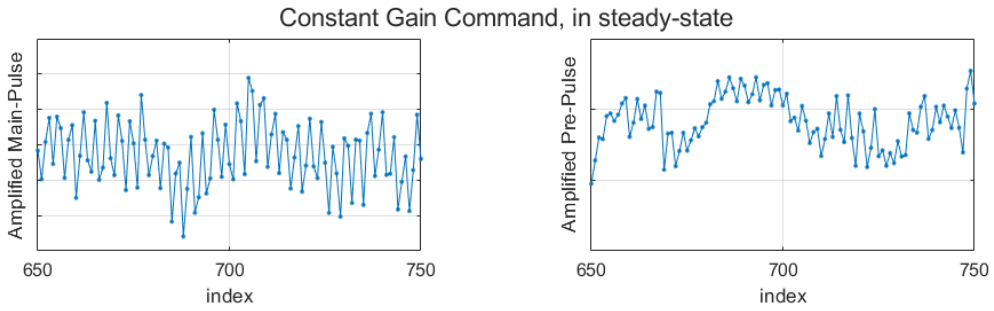


Figure 2.4. Example of zoomed-in amplified laser pulse energy time series, showing oscillation in laser pulse energy. X-axis is index, relative to Gain Command step at index 0. The time range shown corresponds to the steady-state after the Gain Command step change

Besides the second-order transient and the variations trace-to-trace, the individual traces of the pulse energies, at constant Gain Command, also show index-to-index variability. Zooming into the steady-state period, the energy traces display dynamics at various frequencies. The amplified MP energy trace exhibits the most pronounced variation due to an jitter at a high-frequency. The amplified PP energy also shows this disturbance, along with slower, gradual disturbances.

From the time-domain investigation into the experimental data results, we have identified two relevant dynamic behaviors in the laser amplification system. The second-order transient response, occurring over the milliseconds timescale, is a relatively slow dynamic. This dynamic response, resulting from the Gain Command input actuation, can be regulated by appropriate feedback control of the input signal. The high-frequency jitter in the timing signal, however, operates on a much faster timescale than the controller can address. Within the system framework thus far described, there is no appropriate control mechanism that can regulate this behavior. As such, this high-frequency disturbance will be treated as exogenous noise for the purposes of modeling and control.

Chapter 3

Linear time-invariant systems

To develop a model of the system at hand, first the framework of the modeling process must be explained. The process that is under study is that of a single-input-single-output (SISO) system, in our case the input being the Gain Command, and the output being the amplified Main-Pulse energy exiting the drive laser. The inputs and outputs were measured over a period of time during the experiment outlined in the previous section. From the experimental measurements and an understanding of the physical processes underlying the system, a model is developed using prediction error 2-norm minimization methods with the objective of control of the system output.

3.1 State-space descriptions of LTI systems

3.1.1 Continuous time state-space systems

A fundamental description for an input-output system that will be used is a linear time-invariant (LTI) dynamical system. A continuous time LTI dynamical system is represented as a state-space equation

$$\dot{x}(t) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t) + Du(t)$$

where the state $x(t)$ is the internal process of the system, the evolution $\dot{x}(t)$ of which is represented as an ordinary differential equation composed of the linear combination of the homogeneous response $Ax(t)$ and the input response $Bu(t)$. The homogeneous response dictates how the state evolves intrinsically, and the input response dictates how the state evolves based on external forcing. The output $y(t)$ similarly is a linear combination of the state and the input at a given time. The "time-invariance" of the system refers to that the A, B, C, and D matrices do not evolve with time.

The solution to the continuous time linear time-invariant system equation, expressing the trajectory of the state as a function of time, beginning at initial condition $x(t_0)$ at time $t = t_0$, is

$$x(t) = e^{A(t-t_0)}x(t_0) + \int_{t_0}^{t} e^{A(t-\tau)}Bu(\tau)d\tau$$

3.1.2 Discrete time state-space systems

Physical systems, i.e. systems operating in the real world, evolve according to continuous time (CT) dynamics, however computer-based representations of systems are in discrete time (DT). Discrete time representation is used for our modeling purposes as the laser pulses and their measured energies are discrete events. Discrete time representation is also appropriate in the use of sampled data, which is typically the case in computer-controlled systems.

The steps to translate between physical, continuous time dynamics to a discretized computer-based representation are as follows. Rather than using the continuous time variable t, the inputs and outputs are sampled at time instants $t = t_k, k \in \mathbb{Z}^+$ with a sampling interval T_s. The continuous-time input signal $u(t)$ is taken to be a piece-wise continuous signal that maintains a constant value u_k over the interval $T_s = t_{k+1} - t_k$. This constraint on the input signal is that of a zero-order hold (ZOH), akin to the behavior of a digital-to-analog converter:

$$u(t) = u_k \quad \forall\, t \in [t_k, t_{k+1})$$

The evolution of the state $x(t)$ over the interval $t \in [t_k, t_{k+1})$, from an initial condition $x(t_k)$ can then be solved:

$$x(t_{k+1}) = e^{A(t_{k+1}-t_k)}x(t_k) + \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)}Bu_k d\tau$$

Defining $T_s = t_{k+1} - t_k$ and using a change of variables $\tau' = t - \tau$:

$$x(t_{k+1}) = e^{AT_s}x(t_k) + \int_0^{T_s} e^{A\tau'}d\tau' Bu_k$$

The state is then sampled at the $t = t_{k+1}$ instant, $x(t_{k+1}) = x_{k+1}$

$$x_{k+1} = A_d x_k + B_d u_k$$
$$A_d = e^{AT_s}, \quad B_d = \int_0^{T_s} e^{A\tau'}d\tau' B$$

The conversion from the continuous time $y(t)$ to the discrete time y_k is more straightforward, as the matrices $C = C_d$ and $D = D_d$ remain unchanged. The resultant discrete time state space system is thus analogous to the continuous-time physical system, with the change of the matrix transformations.

$$x_{k+1} = A_d x_k + B_d u_k$$
$$y_k = C_d x_k + D_d u_k$$

3.2 Transfer function descriptions of LTI systems

The relationship between the inputs $u(t)$ and output measurements $y(t)$ of the discrete-time linear time-invariant system can be expressed directly via a transfer function. In this representation, the system is defined in discrete time, with the time variable t taken to represent

discrete time steps sampled at time instances $t = t_k,\ k \in \mathbb{Z}^+$

$$y(t) = G_0(q)u(t) \quad t = t_k,\ k \in \mathbb{Z}^+$$

The variable in the transfer function is the time-shift operator q. The q operator operates on a discrete-time signal by shifting the time indices by the exponent on the operator:

$$q^n u(t) = u(t+n) \quad \text{and} \quad q^{-n}u(t) = u(t-n)$$

The transfer function $G_0(q)$ is expressed as a rational function of the operator q. A rational function is the ratio of polynomial functions of orders n_b and n_a, giving $G_0(q)$ the form

$$G_0(q) = \frac{B(q)}{A(q)} = \frac{b_0 + b_1 q^{-1} + ... + b_{n_b-1}q^{-(n_b-1)} + b_{n_b}q^{-n_b}}{1 + a_1 q^{-1} + ... + a_{n_a-1}q^{-(n_a-1)} + a_{n_a}q^{-n_a}}$$

where the denominator polynomial function $A(q)$ is monic. The input-output relationship is then expressed

$$y(t) = G_0(q)u(t) = \frac{B(q)}{A(q)}u(t) =$$

$$b_0u(t) + b_1u(t-1) + ... + b_{n_b-1}u(t-(n_b-1)) + b_{n_b}u(t-n_b)...$$

$$-a_1y(t-1) + ... - a_{n_a-1}y(t-(n_a-1)) - a_{n_a}y(t-n_a)$$

From this expression, the output measurement of the system $y(t)$ is now expressed as a function of the current and past inputs $u(t-n), n = 0 \ldots n_b$ and the past outputs (also known as the initial conditions) $y(t-m), m = 1 \ldots n_a$.

3.2.1 State-space to transfer function conversion

The relationship between the transfer function representation and the LTI state-space difference equation representation of a system can be expressed by taking the z-transform of the

state-space equation. The z-transform converts a discrete time signal (expressed with time index $k \in \mathbb{Z}^+$, corresponding to sampling moments $t = t_k$) into a frequency-domain representation in $z \in \mathbb{C}$. The z-transform of a discrete time signal $x(k)$ is

$$\mathcal{Z}\{x(k)\} = X(z) = \sum_{k=-\infty}^{\infty} x(k) z^{-k}, \text{ expressed as } x(k) \xrightarrow{z} X(z)$$

The z-transform of a signal time-shifted by the operator q^n has a leading factor z^n,

$$x(k+n) = q^n x(k) \xrightarrow{z} z^n X(z)$$

$$\mathcal{Z}\{x(k+n)\} = \sum_{k=-\infty}^{\infty} x(k+n) z^{-k-n+n} = \sum_{k=-\infty}^{\infty} x(k+n) z^{-(k+n)} z^n = z^n X(z)$$

also note that the summation bounds are the same whether the index is k or $k+n$.

With the z-transform defined, the transfer function $G_0(q)$ can be derived

$$\begin{aligned} x_{k+1} &= A_d x_k + B_d u_k \\ y_k &= C_d x_k + D_d u_k \\ &\downarrow z \\ zX(z) &= A_d X(z) + B_d U(z) \\ Y(z) &= C_d X(z) + D_d U(z) \end{aligned}$$

then solving for $y(t)$

$$\begin{aligned} Y(z) &= (C_d(zI - A_d)^{-1} B_d + D_d) U(z) \\ &\downarrow z^{-1} \\ y(t) &= (C_d(qI - A_d)^{-1} B_d + D_d) u(t) \\ G_0(q) &= C_d(qI - A_d)^{-1} B_d + D_d \end{aligned}$$

3.2.2 Deterministic and stochastic parts of systems

The transfer function $G_0(q)$, called the "plant model" of the system, describes the deterministic relationship between the input and output signals of the system. In an idealized system, the output signal would be determined entirely by the plant model and the input signal. In a practical system, however, there will be additional variation in the system output, ostensibly independent of the input signal, which may be attributed to external disturbance and dynamics not captured by the plant model. The extra variation is modeled as additive "noise" and is described as the convolution of the "noise model" (or noise filter) $H_0(q)$ with an input discrete time random process $e(t)$

$$v(t) = H_0(q)e(t)$$

. The random process e(t) has mean $\mathcal{E}\{e(t)\} = 0$, variance $\mathcal{E}\{e^2(t)\} = \lambda$, and autocorrelation $\mathcal{R}\{e(t,\tau)\} = \mathcal{E}\{e(t)e(t-\tau)\} = 0 \quad \forall \tau \neq 0$, and is uncorrelated with the known input $u(t)$: $e(t) \perp u(t)$ [11].

The complete transfer function description of the data-generating system is the additive sum of the deterministic input signal, filtered by the plant model, and the random process, filtered by the noise model:

$$\mathcal{S} : y(t) = G_0(q)u(t) + H_0(q)e(t)$$

This is the relationship that we intend to capture via system modeling.

Chapter 4

System identification

4.1 System identification introduction

The objective of modeling is to use the measured input and output data taken during the experiment to estimate the internal structure of the system; in our case, the plant and noise models of the system $\mathcal{S} : y(t) = G_0(q)u(t) + H_0(q)e(t)$. With a sufficiently accurate model of the sytem we can predict the impact of future inputs into the system. The model need not be a mirror image facsimile of the physical system in order to be useful; the measurements of a high-order physical system can be closely approximated by a low-order model. The process of estimating a model for a system is called "system identification." This practice relies on a set of fundamental tools, generally outlined by Lennart Ljung [12] and implemented in the MATLAB System Identification Toolbox.

The objective of system identification is to determine an approximation of the system $\mathcal{S}$, notated as the model $\mathcal{M}$,

$$\mathcal{M} : y(t) = G_\theta(q)u(t) + H_\theta(q)e(t)$$

The model $\mathcal{M}$ is composed of transfer functions G_θ and H_θ, which are rational functions in q, and $e(t)$ is white noise in the model.

The transfer functions can then be parametrized (as indicated by the subscript θ) by the

leading coefficients b_i, f_i, c_i, and d_i of the terms in the numerator and denominator polynomials. In parametric system identification, the objective is to select this set of parameters such that the predicted output of the model $\mathcal{M}$ matches as closely as possible to the measured output of the data-generating system $\mathcal{S}$.

$$\mathcal{M}: \quad y(t) = G_\theta(q)u(t) + H_\theta(q)e(t) = \frac{B_\theta}{F_\theta}u(t) + \frac{C_\theta}{D_\theta}e(t) =$$
$$\frac{b_0 + b_1 q^{-1} + ... + b_{n_b} q^{-(n_b)}}{1 + f_1 q^{-1} + ... f_{n_f} q^{-n_f}} u(t) + \frac{1 + c_1 q^{-1} + ... + c_{n_c} q^{-n_c}}{1 + d_1 q^{-1} + ... d_{n_d} q^{-n_d}} e(t)$$

There are three general steps to use system identification to arrive at a suitable model. First, the model structure must be decided. There are numerous model structures: ARX, FIR, ARMAX, OE, Box-Jenkins, etc. These are categorized based on which of the B_θ, F_θ, C_θ, and D_θ polynomials are set equal to 1, and whether the denominator F_θ and D_θ polynomials share common dynamics. Depending on the selection of the structure, the parametric optimization steps will differ. The MATLAB System Identification Toolbox has built-in functions to perform parametric optimization using these structures, and it is up to the user to select which structure is appropriate.

Secondly, the model order for each polynomial must be decided. The model order refers to the number of terms in each polynomial: n_b, n_f, n_c, and n_d. A higher-order model may capture more of the dynamics in the data-generating system, but will be more computationally intensive to perform the parametric optimization. An overly-parametrized model can result if the selected model order is too large given the system dynamics, in which case the higher-order terms do not contribute to the model accuracy and a model with fewer parameters would suffice. Selecting too-low a model order, however, risks under-fitting the dynamics of interest. An appropriate model order can be analytically determined using functions such as Akaike's Information Criterion (AIC) or Rissanen's Minimum Description Length (MDL) [12]. Otherwise, the model order for

each of the polynomials can be selected directly by the user, and are input arguments into the MATLAB System Identification Toolbox functions.

Lastly, the parameters b_i, f_i, c_i, and d_i must be determined. This step is accomplished using parametric optimization algorithms in the MATLAB Sytem Identification Toolbox functions.

4.2 Prediction error 2-norm minimization

4.2.1 Prediction error formulation

Once a candidate model structure, order, and parametrization θ have been decided on, the model output is compared to the data-generating system to quantify how well the model performs. Candidate parametrizations may be iterated through in order to optimize the model fit to the data-generating system and determine the optimal parametrization $\hat{\theta}$. The model performance is measured via the prediction error $\varepsilon(t,\theta)$

$$\varepsilon(t,\theta) = y(t) - y(t|t-1,\theta)$$

where $y(t|t-1,\theta)$ is the prediction of the measurement at time t, given measurements $y(\tau)$, $\tau \leq t-1$, and a model parametrization θ.The model performance can be quantified by taking the 2-norm of the prediction error $\varepsilon(t,\theta)$ over time

$$\|\varepsilon(t,\theta)\|_2 = \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t,\theta)^2$$

The optimal parametrization $\hat{\theta}$, and therefore the optimal model of a given structure and order, is therefore the model that minimizes the prediction error 2-norm over the parameter space.

$$\hat{\theta} = \min_{\theta} \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t,\theta)^2$$

To understand this quantification, we begin by defining the model prediction error $\varepsilon(t,\theta)$. To begin, the noise component $v(t) = H_0(q)e(t)$ of the data generating system $\mathcal{S}$ is written to include the next-step ahead prediction of the noise term:

$$v(t) = H_0(q)e(t) = (1 + h_1 q^{-1} + h_2 q^{-2} + ...)e(t) = (1 + \bar{H}_0(q))e(t) = e(t) + v(t|t-1)$$

where $\bar{H}_0(q) = H_0(q) - 1$ has a one-step time delay (as the leading term is in q^{-1}).

The one-step-ahead noise prediction is

$$v(t|t-1) = \bar{H}_0(q)e(t) = (H_0(q) - 1)H_0(q)^{-1}v(t) = (1 - H_0(q)^{-1})v(t)$$

The one-step-ahead noise prediction represents the best estimate of the upcoming noise contribution, given past measurements.

The one-step-ahead output prediction can then be formed by combining the known deterministic input response with the one-step-ahead noise prediction

$$\begin{aligned} y(t|t-1) &= G_0(q)u(t) + (1 - H_0(q)^{-1})v(t) \\ &= G_0(q)u(t) + (1 - H_0(q)^{-1})(y(t) - G_0(q)u(t)) \\ &= H_0(q)^{-1}G_0(q)u(t) + (1 - H_0(q)^{-1})y(t) \end{aligned}$$

As we do not have the full information of the transfer functions of the physical system, the one-step-ahead prediction of the output is written using the model transfer functions

$$y(t|t-1,\theta) = H_\theta(q)^{-1}G_\theta(q)u(t) + (1 - H_\theta(q)^{-1})y(t)$$

The prediction error can now be defined

$$\begin{aligned}\varepsilon(t,\theta) &= y(t) - y(t|t-1,\theta) = H_\theta^{-1}(q)(y(t) - G_\theta(q)u(t)) \\ &= H_\theta(q)^{-1}(G_0(q)u(t) - G_\theta(q)u(t) + H_0(q)e(t) - H_\theta(q)e(t)) + e(t)\end{aligned}$$

4.2.2 Parseval's theorem

The minimization problem can be restated using Parseval's theorm [14]:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \varepsilon(t)^2 = \frac{1}{2\pi} \int_{\omega=-\pi}^{\pi} \phi_\varepsilon(\omega) d\omega \tag{4.1}$$

This theorem equates the 2-norm of the time series of the signal $\varepsilon(t)$ (as the number of samples $N \to \infty$) to the integral of the signal's spectrum $\phi_\varepsilon(\omega,\theta) = \mathcal{F}(R_\varepsilon(\tau))$, $R_\varepsilon(\tau)$ being the autocorrelation of the signal and $\mathcal{F}$ the discrete time Fourier transform (DTFT). In our particular case the signal is the prediction error $\varepsilon(t,\theta)$.

Per Ljung, Theorem 2.2 [12], the spectrum of the prediction error can be completely expressed

$$\phi_\varepsilon(\omega,\theta) = |\frac{G_0(\omega) - G_\theta(\omega)}{H_\theta(\omega)}|^2 \phi_u(\omega) + |\frac{H_0(\omega) - H_\theta(\omega)}{H_\theta(\omega)}|^2 \phi_e(\omega) + \phi_e(\omega)$$

where $\phi_e(\omega) = \lambda$ is the constant spectrum of the white noise input $e(t)$, and recall that $e(t)$ is zero mean and $u(t)$ and $e(t)$ are independent.

The prediction error 2-norm minimization problem can be restated

$$\hat{\theta} = \min_\theta \frac{1}{2\pi} \int_{\omega=-\pi}^{\pi} |\frac{G_0(\omega) - G_\theta(\omega)}{H_\theta(\omega)}|^2 \phi_u(\omega) + |\frac{H_0(\omega) - H_\theta(\omega)}{H_\theta(\omega)}|^2 \lambda \quad d\omega$$

The parametrization is globally minimized in the case that the plant model and noise model perfectly mirror the physical plant and noise dynamics, bringing the numerator terms to zero. However, this is not practical in the real system. As there will necessarily be some error in the

modeling, the m inimization can be influence by the input signal spectra $\phi_u(\omega)$ and the candidate noise model $H_\theta(\omega)$.

4.3 Model structures

Depending on the chosen model structure, the process of optimizing the parametrization such that the prediction error 2-norm is minimized will be different. Some model structures, such as the ARX model, have explicit solutions to the prediction error 2-norm minimization problem, while other structures will rely on iterative methods to perform the minimization.

4.3.1 ARX model structure

An ARX (Auto-Regressive with eXogenous input) model is of the form

$$\mathcal{M}_{ARX}: \quad y(t) = \frac{B_\theta}{A_\theta}u(t) + \frac{1}{A_\theta}e(t)$$

also note that the denominator polynomials are equal. This structure provides the ARX model framework an interesting result: the parameter selection that minimizes the summed squared

prediction error can be computed explicitly using linear regression.

$$\begin{aligned} y(t) =& b_0u(t) + b_1u(t-1)... + b_{n_b-1}u(t-(n_b-1)) + b_{n_b}u(t-n_b)... \\ & - a_1y(t-1)... - a_{n_a-1}y(t-(n_a-1)) - a_{n_a}y(t-n_a) + e(t) \\ = & \begin{bmatrix} u(t) & u(t-1) & ... & u(t-n_b) & -y(t-1) & ... & -y(t-n_a) \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{n_b} \\ a_1 \\ \vdots \\ a_{n_a} \end{bmatrix} + e(t) \\ = & \phi^T(t)\theta + e(t) \end{aligned}$$

where $\phi^T(t)$ is the regressor vector, θ is the parameter vector, and $e(t)$ is white noise in the model.

The one-step-ahead prediction of the model is $y(t|t-1,\theta) = \phi^T(t)\theta$, as the noise term is white (and therefore uncorrelated with previous inputs or outputs). The prediction error is then be defined

$$\varepsilon(t,\theta) = y(t) - \phi^T(t)\theta$$

By stacking the series of output measurements $y(t)$ into a vector Y, the series of prediction errors $\varepsilon(t,\theta)$ into a vector ε, and the series of regressor vectors into a matrix Φ, we can rewrite

$$\varepsilon = Y - \Phi\theta$$

and knowing that the optimized prediction error vector $\varepsilon(\hat{\theta},t)$ is uncorrelated to the regressor

vector, $\varepsilon(\hat{\theta},t) \perp \Phi$ (i.e. $\varepsilon(\hat{\theta},t)^T\Phi = 0$), then the parameter vector can be solved directly

$$\hat{\theta} = [\Phi^T\Phi]^{-1}\Phi^T Y$$

4.3.2 OE model structure

An Output-Error (OE) model structure is defined as having the form

$$\mathcal{M}_{OE}: \quad y(t) = \frac{B_\theta}{F_\theta}u(t) + e(t)$$

The OE model structure is advantageous as the noise model is unity, meaning that the noise contribution to the output is white noise without dynamics. This structure may be selected in case it is presumed that the noise and the plant model do not share any common dynamics, and the noise is essentially white.

This optimized parameter vector in the OE model structure

$$\hat{\theta} = \begin{bmatrix} \hat{b_0} & \hat{b_1} & \ldots & \hat{b_{n_b}} & \hat{f_1} & \ldots & \hat{f_{n_f}} \end{bmatrix}^T$$

cannot be solved explicitly using the same linear-regression method as the ARX model structure. The identification is accomplished using a non-linear Gauss-Newton least-squares search.

4.3.3 Box-Jenkins model structure

A Box-Jenkins model structure has the form

$$\mathcal{M}_{BJ}: \quad y(t) = \frac{B_\theta}{F_\theta}u(t) + \frac{C_\theta}{D_\theta}e(t)$$

In Box-Jenkins the noise model may have dynamics distinct from the plant system, each of which is independently parametrized. While more complicated, the modeling capabilities are greater. Similar as to an OE model structure, the parameter $\hat{\theta}$ is solved using a non-linear search.

Chapter 5

System analysis and modeling results

The experiments carried out on the high-power drive laser system, outlined in Chapter 2, produced input-output measurements that can be used to identify the underlying system dynamics via the system identification framework. The experimental data analyzed is of the input Gain Command signal step function, while the output amplified Main-Pulse laser energies are recorded at the exit of the amplification stage. A single cycle of the step is used for system identification, with the data windowed to a 300 index period centered at the step moment.

The input-output data from the experiment, along with the selected model order (n_a, n_b ,n_f) are input into the selected MATLAB system identification function, which returns the identified model parametrization $\hat{\theta}$. The resultant model parametrization is used to simulate the system response to the step input $y_{sim}(t) = G_{\hat{\theta}}(q)u(t)$. The simulation and the measured output are then overlaid to compare the model results. Using the open-loop simulation result as a selection heuristic differs from the model fitting criterion, which uses the one-step-ahead predictions. The open-loop simulation behavior is better-suited to our eventual control design objective.

5.1 Analysis of the high-frequency disturbance

The laser system exhibited two relevant dynamic behaviors: the second-order transient effect (milliseconds timescale) and the high-frequency jitter. The high-frequency jitter was

found to be due to an external disturbance. The specific frequency of the external disturbance is found to appear in the amplified Main-Pulse and Pre-Pulse signals. This frequency is labelled as ω_0. Additionally, as shown in 2.4, the high-frequency jitter is apparent when the input Gain Command is at steady-state. These observations help to conclude that the high-frequency jitter can be characterized as an exogenous disturbance that should be treated as separate from the amplification process.

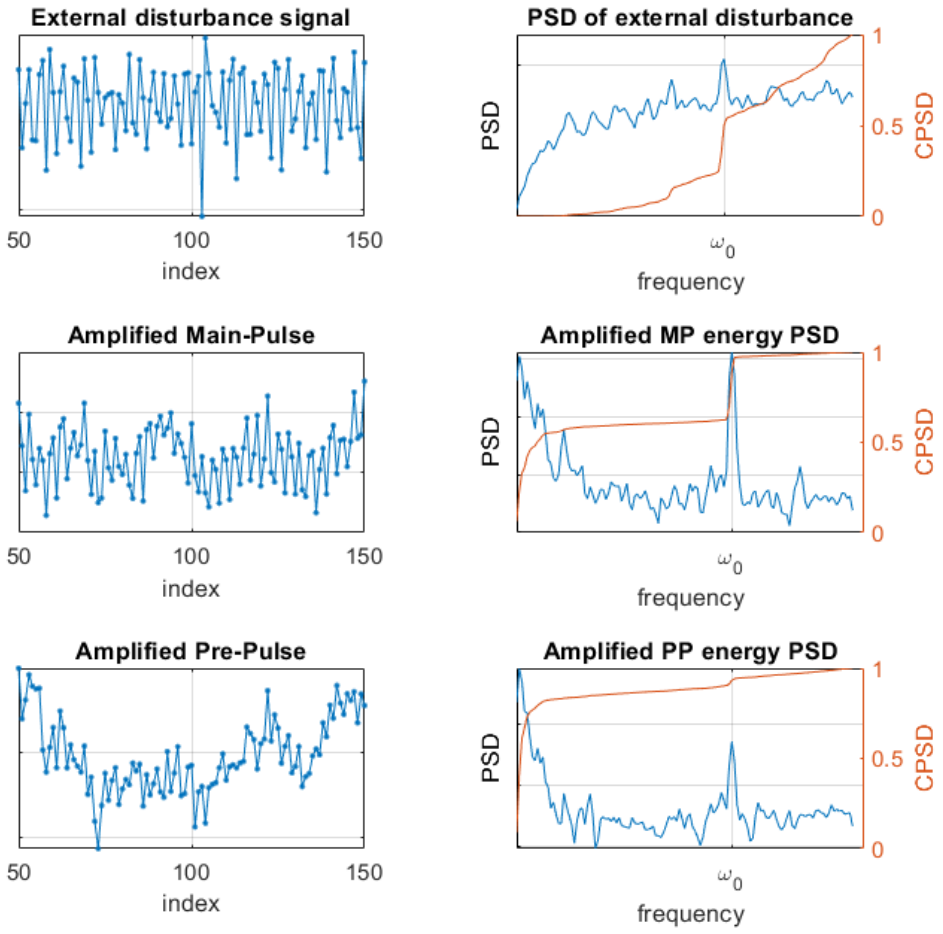


Figure 5.1. Time series and Power Spectral Density (PSD) plots of the external disturbance and amplified outputs for the Main-Pulse and Pre-Pulse. The PSD plots show the Power Spectral Density on the left y-axis, and the normalized cumulative PSD on the right y-axis. The external disturbance displays an oscillation at ω_0, which cascades to the amplifier output

The high-frequency jitter, being due to a known exogenous disturbance to the system, can be addressed in the modeling process by constraining the exogenous mode to the noise model. In this way, the plant model dynamic that predicts the amplified laser energy as a product of the amplifier Gain Command input can be determined via system identification.

5.2 Data filtering and model identification

The behavior we aim to capture in the identified model is the second-order transient response in the windowed period after the commanded step. This transient response is deemed "second-order" as it is characterized by the oscillation and dampening associated with a second-order linear system with complex conjugate poles. The magnitude of the overshoot and settling time are the characteristics via which it is qualified whether the model is adequately capturing the system dynamics. The high-frequency jitter, while part of the physical system, is not desired to be represented in the identified plant model, and instead should be captured in the noise model.

The capability of the system identification architecture to produce a model with the desired dynamic response behavior is stymied by the high-frequency jitter intrinsic to the output signal (see Section 5.3). The exogenous jitter impacts the capability of the identification framework to capture the dynamics of interest in the identified model. By removing the exogenous spectral content from the input and output signals, the system identification framework can adequately capture the low-frequency second-order transient into the identified model, via minimizing the prediction error 2-norm. As the exogenous noise has specific dynamics, the impact can be nullified by filtering the data with a specific filter.

Consider the generic case of a model structure with independently-parametrized plant and noise models, expressed in a Box-Jenkins model structure.

$$y(t) = \frac{B_\theta(q)}{F_\theta(q)}u(t) + \frac{C_\theta(q)}{D_\theta(q)}e(t)$$

In our case, the noise model $\frac{C_\theta(q)}{D_\theta(q)}$ contains the dynamics that introduce the high-frequency mode.

By filtering both sides by the inverse of the noise model $\frac{D_\theta(q)}{C_\theta(q)}$, the system can be rewritten into an Output Error model structure, as the noise dynamics have been nullified.

$$\begin{aligned} \frac{D_\theta(q)}{C_\theta(q)}y(t) &= \frac{B_\theta(q)}{F_\theta(q)}\frac{D_\theta(q)}{C_\theta(q)}u(t) + e(t) \\ y_f(t) &= \frac{B_\theta(q)}{F_\theta(q)}u_f(t) + e(t) \end{aligned} \tag{5.1}$$

The filtered input $\frac{D_\theta(q)}{C_\theta(q)}u(t) = u_f(t)$ and filtered output $\frac{D_\theta(q)}{C_\theta(q)}y(t) = y_f(t)$ have the same plant model $\frac{B_\theta(q)}{F_\theta(q)}$ as the unfiltered signals. This is important as it shows that filtering the data will not impact the validity of the resultant identified model.

Filtering the signals by $\frac{D_\theta(q)}{C_\theta(q)}$ is achieved by emulating the inverse noise model dynamics. With the assumption that the noise model introduces the ω_0 mode, a filter is defined that attenuates the exogenous high-frequency mode. A 20th-order bandstop Chebyshev filter is selected, with bandstop range windowing the high-frequency mode at ω_0.

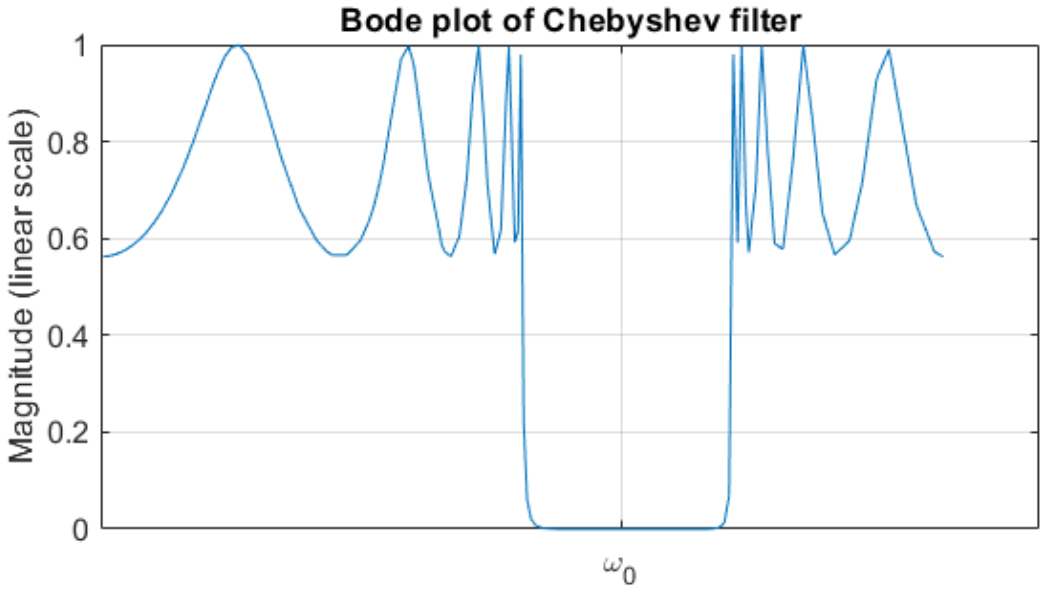


Figure 5.2. Bode plots of the Chebyshev filter. The filter attenuates the spectrum in a window around the exogenous disturbance mode ω_0

The spectrum of the original and the filtered signals, along with their time series, are

shown below. The filtered output time series has lower variation, indicating that the noisy appearance of the raw output time series is due to the ω_0 mode. Note that the baseline level of the spectral content between the windowed bandstop region is offset between the filtered output and filtered input. This will be relevant for model identification as the model transfer function may amplify spectral content in this region to account for the different baseline levels.

The Chebyshev bandstop filter has a -5 dB gain at low frequencies (below 5kHz). The non-zero-dB DC gain is due to the 5 dB passband ripple that is selected in the construction of the filter when using the MATLAB cheby1 function.

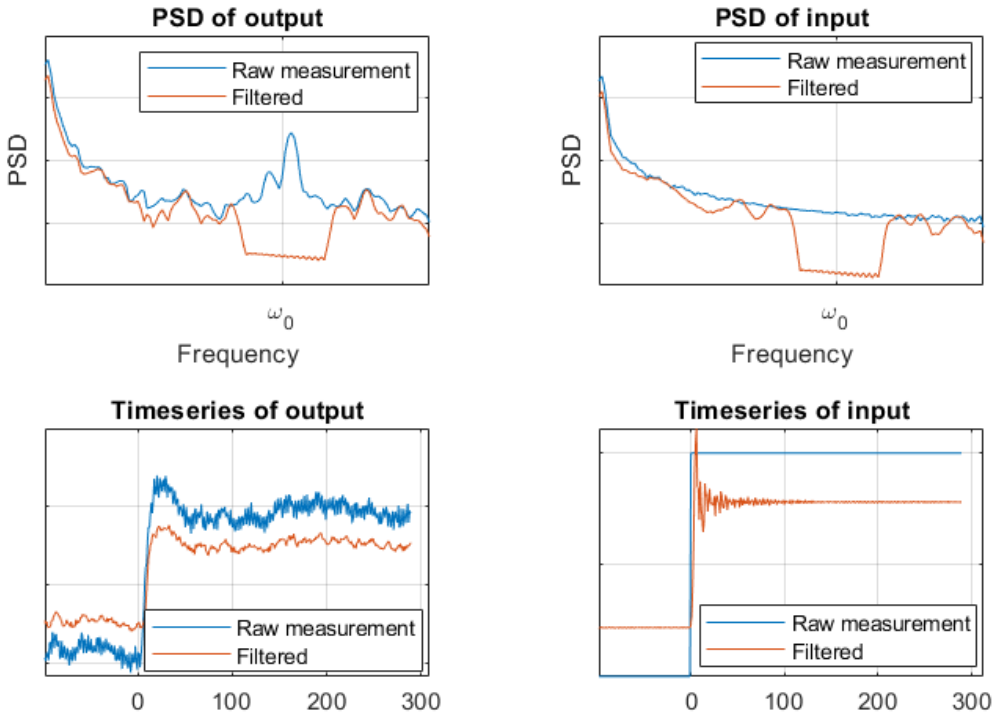
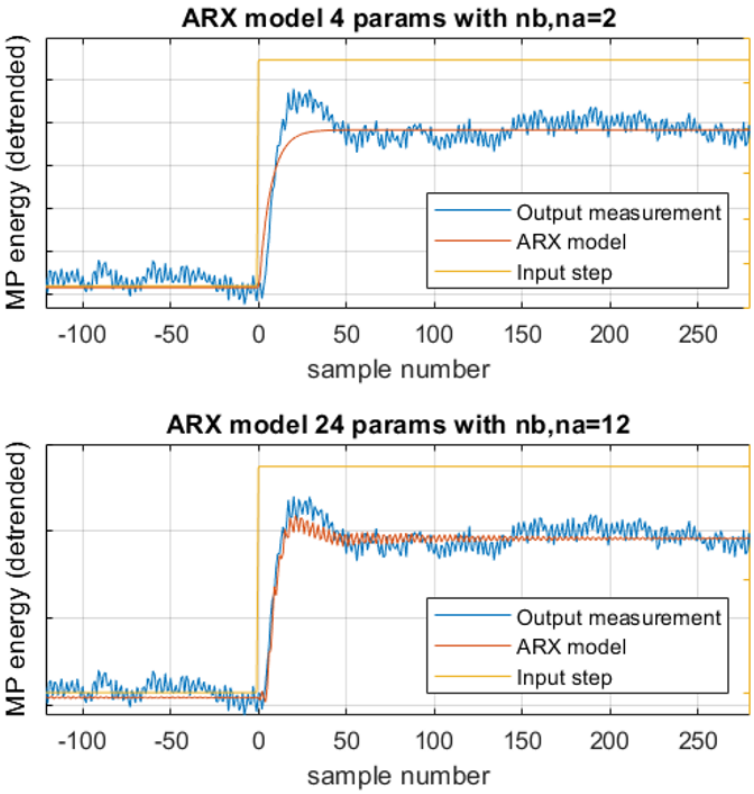


Figure 5.3. The effect of the high-frequency mode-attenuating Chebyshev filter. The spectral content of the input and output signals is shown to be dramatically attenuated within the bandstop region centered around ω_0

5.3 Modeling results without filtering applied

In order to demonstrate the efficacy of prefiltering the data, we first consider the case of identifying models from the raw data without applying filtering. Three different models, using ARX and OE structures of varying model orders, were identified. As shown in the figures below, the raw data collected from the experiment, without filtering applied, produced identified models that are either under-parametrized (top figure) or that amplify the ω_0 mode considerably (middle and bottom figures), without appropriately capturing the second-order transient effect.



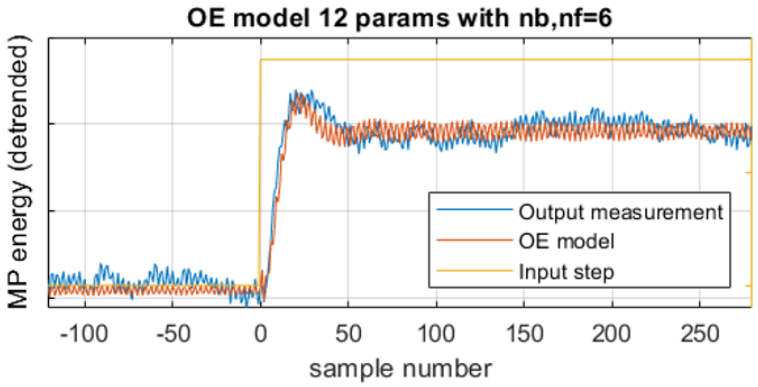


Figure 5.4. System modeling results using unfiltered data, showing the step response of the identified model against the measured response. The simulated outputs are either under-parameterized or amplify the ω_0 mode, contrary to our modeling objective

The model performance can be quantified using Parseval's theorem, as outlined in Chapter 4. The prediction error is measured using a 1-step ahead predictor based on the identified model, and the prediction error is $\varepsilon(t,\hat{\theta}) = y(t) - y(t|t-1,\hat{\theta})$. The spectrum of the prediction error is calculated using Welch's method on the prediction error time series. In the three model structures tested, the spectral content shows a substantial peak at ω_0, this indicates that the prediction error 2-norm minimization must prioritize fitting this behavior, whereas the desired behavior is for a good fit at the low-frequency end of the spectrum.

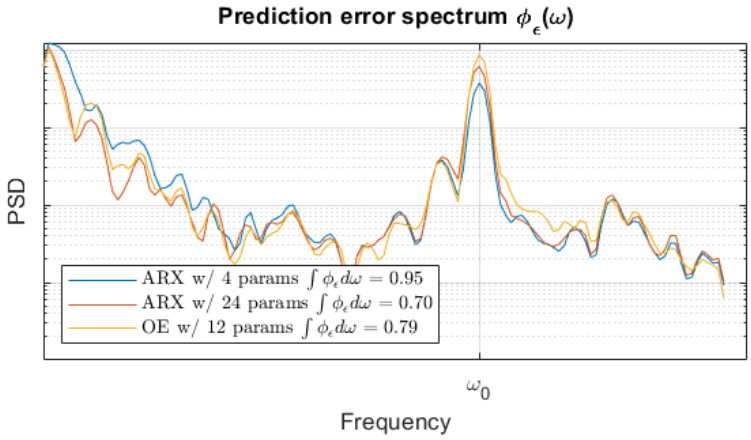


Figure 5.5. System modeling results, showing the prediction error spectrum of the three candidate models. Each of the models shows substantial prediction error spectral content around ω_0

5.4 Modeling results on filtered data

The filtered input $u_f(t)$ and output $y_f(t)$ data signals, defined in Section 5.2, provide the foundation upon which to construct a model that fits our objectives, namely to simulate the overshoot and settling dynamic of the laser system in the milliseconds timescale while ignoring the modeling of the ω_0 component of the signal, at this stage. This modeling is accomplished using an Output Error model, as explained in Equation (5.1). The model parametrization is performed using the MATLAB System Identification Toolbox oe function, which performs a non-linear Gauss-Newton least-squares optimization to arrive at a parametrization that minimizes the 2-norm of the prediction error. The Output Error structure is chosen because the filtering captures the known ω_0 noise model. The identification is done with the default MATLAB oe function settings.

The remaining step for us is to decide on the model order. In the Output Error model structure, the model order consists of the polynomial orders of the numerator and denominator plant model transfer functions $B_\theta(q)$ of order n_b and $F_\theta(q)$ of order n_f, and the number of delays between the input and output n_k. The candidate set of model structure parameters selected by the user is input, along with the filtered input and output data, in the OE parameter optimization

algorithm to arrive at a candidate model.

After careful consideration, the selected model structure is of the structure $n_b = 7$, $n_f = 7$, and $n_k = 0$.

The transfer function of the identified model is:

$$y(t) = \frac{B(q)}{F(q)}u(t) + e(t) \tag{5.2}$$

$$B(q) = 0.008181 + 0.01201q^{-1} - 0.1232q^{-2} + 0.1782q^{-3} + ...$$
$$0.178q^{-4} - 0.1573q^{-5} - 0.06625q^{-6}$$

$$F(q) = 1 - 0.7692q^{-1} - 1.799q^{-2} + 1.212q^{-3} + 1.629q^{-4} + ...$$
$$-1.131q^{-5} - 0.5952q^{-6} + 0.4697z^{-7}$$

The model performance, and therefore the quality of the candidate model structure, was validated on the following criteria:

5.4.1 Stability

The transfer function must be stable; in discrete time, this requires that the poles of the transfer function lie within the unit disk. The Output Error algorithm intentionally will only produce stable models, as the fitting involves simulation of the system. As such, the candidate model is stable, with all poles having magnitude $|p| < 1$. This is shown in the pole-zero plot below. Note also that there are zeros that lie outside the unit disk.

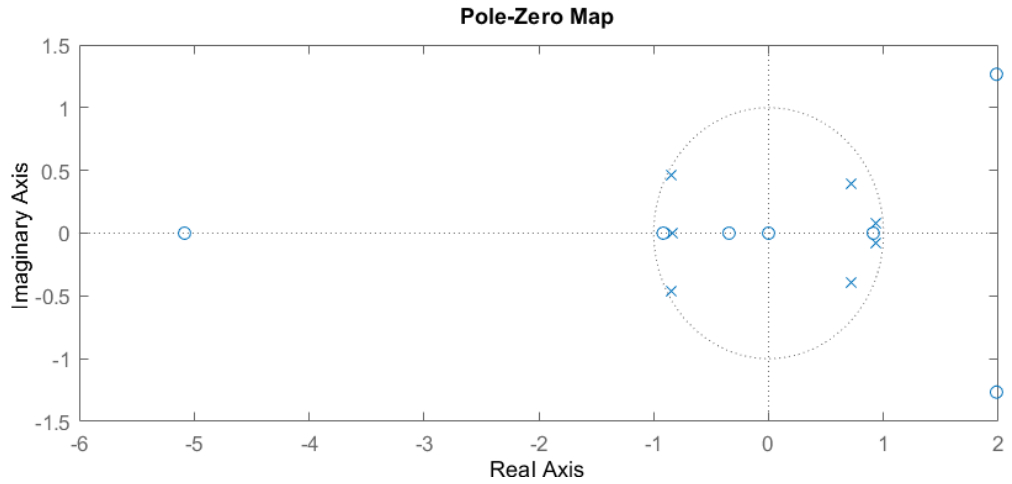


Figure 5.6. Pole-zero plot of the 14-parameter Output Error model

5.4.2 Non-singularity

As will be explained later, the transfer function representation of the model must be invertible, this allows the system inverse to be used for open-loop control. In a state-space representation, this is achieved by having a non-zero feed-forward matrix D ($y(t) = Cx(t) + Du(t), D \neq 0$). In the transfer function representation, this is prescribed by having no delays in the system, $n_k = 0$. A system with a delay between the input and the output, i.e. $n_k \neq 0$, cannot be inverted without violating causality.

While the identified model is invertible, the inverse is not stable, due to the zeros outside the unit circle (see Figure 5.6). The instability poses challenges for open-loop control. The implications of this instability are explored further in Section 6.2.

5.4.3 Overshoot and settling dynamics of time series

The simulated response of the 14-parameter Output Error model (5.2) to the filtered input closely matches the measurement of the filtered output data. The magnitude and settling time of the modeled response are well-aligned to the measurement, and the modeled response stabilizes to the steady-state value within 100 samples after the initial change in the filtered input. The ω_0

exogenous disturbance is not present in the steady-state model response.

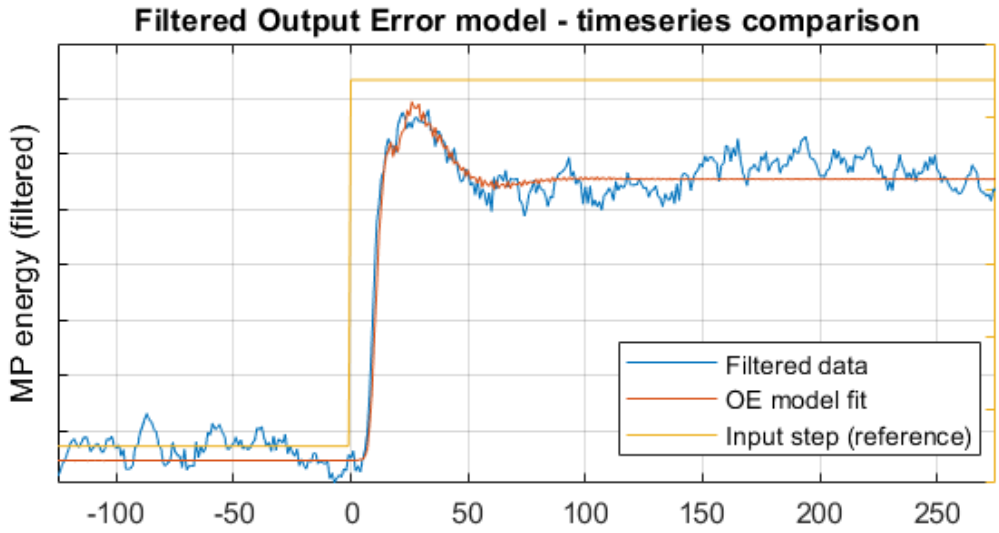


Figure 5.7. Time series of the filtered ouput data and the simulated response of the OE model to the filtered input, parametrized on the filtered input and output data

5.4.4 Minimized prediction error 2-norm

The improvement in the model performance using filtered input and output data, compared to the unfiltered data, is quantified by comparing the prediction error spectra. As outlined in Chapter 4.2, the predictive quality of the model is measured by 2-norm of the prediction error across the time series, then calculated as the integral of the spectrum of the prediction error via Parseval's theorem (Equation (4.1)).

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \varepsilon(t)^2 = \frac{1}{2\pi} \int_{\omega=-\pi}^{\pi} \phi_\varepsilon(\omega) d\omega$$

The prediction error is measured using a 1-step ahead predictor, and the spectrum of the prediction error is calculated using Welch's method.

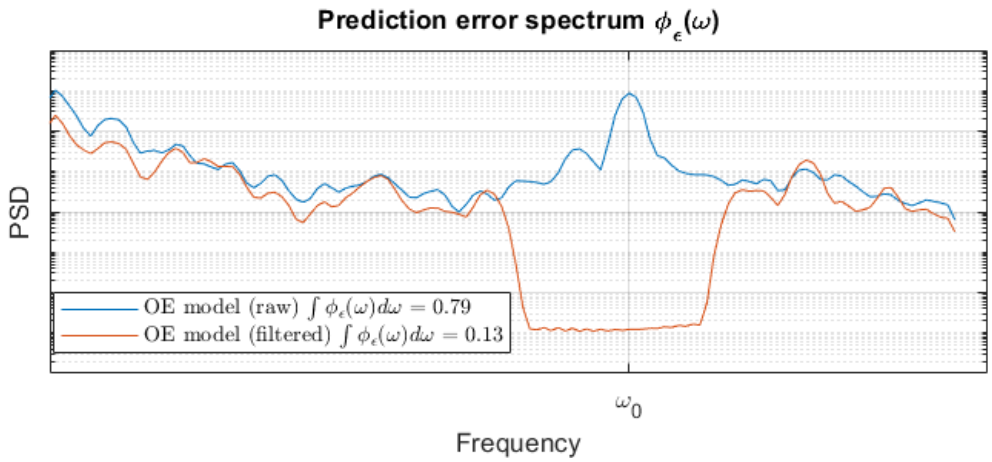


Figure 5.8. The prediction error spectrum of the parametrized models, before and after applying the Chebyshev filter

It is shown that the prediction error spectra of the filtered- and unfiltered-data-based models are similar between outside the Chebyshev filter range. Within the region of the Chebyshev filter bandstop, however, the spectrum of the prediction error of the filtered-data-based model is substantially lower in magnitude. By attenuating the ω_0 mode, the total area under the spectrum of the prediction error is reduced, this indicates that the model parametrized on the filtered data has a better predictive capability.

5.4.5 Transfer function frequency response and the spectral estimate

The transfer function spectral estimate of the filtered data is calculated as the ratio between the cross-spectrum between the filtered input and output time series measurements, and the auto-spectrum of the filtered input time series, both spectra calculated using Welch's method. Interestingly, the transfer function estimate shows amplification in the Chebyshev bandstop range, This is because the filtered input and output have different noise floors that the spectra are attenuated to, as can be seen in Figure 5.3. While the transfer function estimate therefore shows an amplification in this range, there is very little spectral content in either the input or output here.

The spectral estimate is compared to the frequency response of the parametrized model (5.2) ($G_\theta(e^{j\omega})$). It is shown that the low-frequency behavior of the spectral estimate is captured by the model. The dynamics of interest, namely the overshoot and settling within the first milliseconds timescale of the step function response, are well within this range. This further explains the model's validity in capturing the system dynamics. Above the Chebyshev window range, the model frequency response shows a peak at ω_1, in line with the spectral estimate which shows a peak at ω_1.

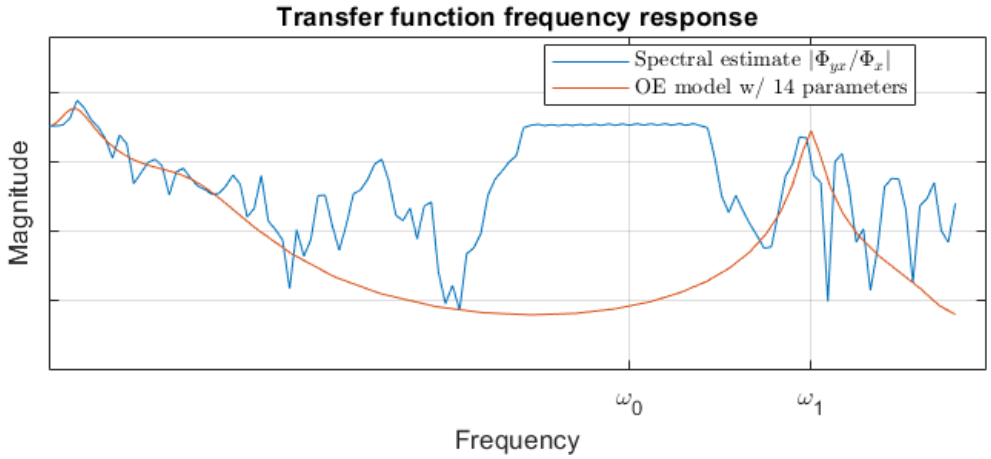


Figure 5.9. The model transfer function frequency response is compared to the spectral estimate, calculated from the time series, showing good agreement in the frequency response at low frequencies

Chapter 6

Model applications

6.1 Model performance on unfiltered data

By pre-filtering the data using the Chebyshev filter, the known noise-model dynamics have been removed from the system, and a successful model (5.2) has been constructed which predicts the low-frequency overshoot dynamic found in the measured data. The model performance is ultimately validated by comparing the step-response of the model, i.e. the simulated response of the model to the unfiltered Gain Command input step, to the measured (unfiltered) output amplified laser pulse energies at the exit of the amplifier chain.

As predicted, the step-response of the model performs well, matching the over-shoot and settling dynamic of the measured data. The ω_0 jitter that is apparent in the measured data crucially does not appear in the model response.

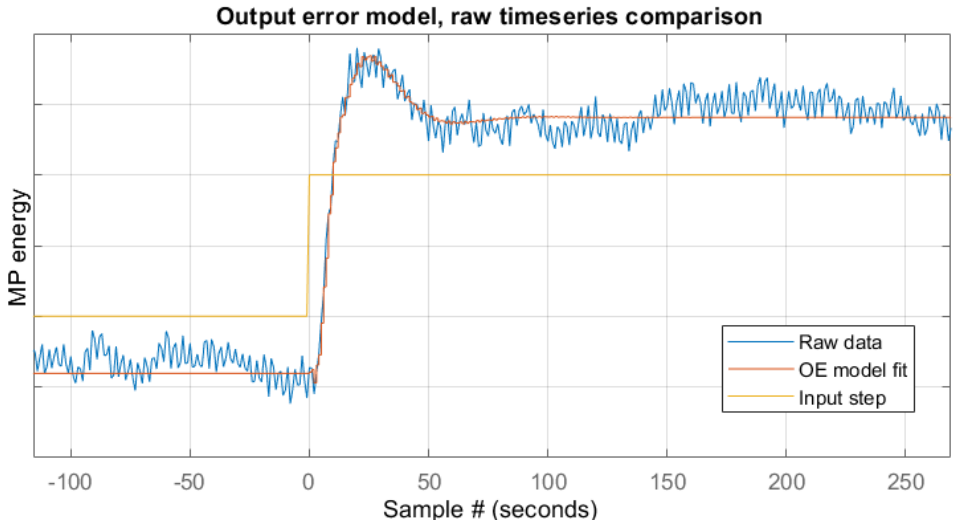


Figure 6.1. Time series validation of the model, showing the overshoot dynamic present in the unfiltered data is captured in the step response of the Output Error model

The simulated response of the model is also shown to fit the ensemble response of the entire time series. The ensemble comparison confirms that the trace used for parametrizing the model was representative of the general response of the system. The simulated response, using the Gain Command step-function as the input, matches the overshoot and steady-state value of the ensemble.

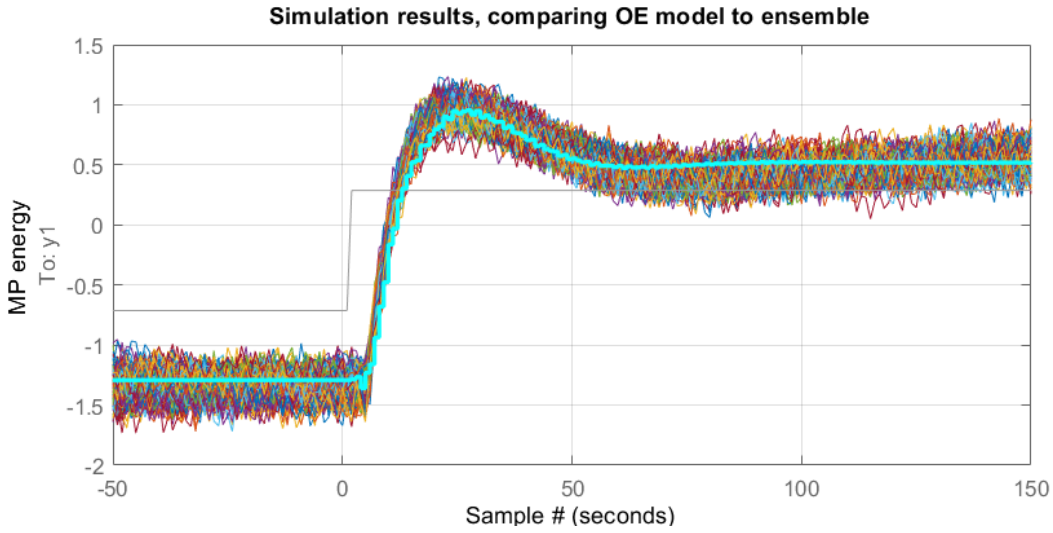


Figure 6.2. Model simulation result (in turquoise) compared to the ensemble response of the time series, each trace corresponding to 50 indices before to 150 indices after the Gain Command step, with 164 traces overlapped

6.2 Open-loop control

The model that has been developed demonstrates good performance when used for simulation and prediction of the high-power laser system. The next step to qualify the capability of the model is to judge its applicability for control. The fundamental control system problem that we will consider is open-loop control via model inversion. This topic has been treated in Hespanha [16], lecture 20, and Dautt-Silva [17], chapter 2.

Consider a reference signal $r(t)$ which is input into the plant model $G_\theta(q)$, in our case the reference signal is the Gain Command signal and the plant is modeled by the Output Error transfer function. The objective is to drive the output $y(t)$ along a desired trajectory $y_d(t)$. The trajectory $y_d(t)$ is a filtered version of the reference signal $r(t)$, $y_d(t) = Q(q)r(t)$; as a trivial example $Q(q) = I$ achieves the goal that the output trajectory $y_d(t)$ exactly tracks the reference input $r(t)$. However given the plant dynamics, $y(t) = G_\theta(q)r(t) \neq y_d(t)$ generally, without additional control applied to the system.

The goal is accomplished by inserting a controller in-between the reference signal and

the plot model to cancel out the plant dynamics and insert the desired dynamics: $C_\theta(q) = G_\theta(q)^{-1}Q(q)$. Therefore if we can define the inverse of the model, the system can be controlled to a desired output trajectory, without the need for, or benefits from, feedback.

$$r(t) \rightarrow \boxed{C_\theta(q) = G_\theta(q)^{-1}Q(q)} \xrightarrow{u(t)} \boxed{G_\theta(q)} \rightarrow y(t)$$

$$y(t) = Q(q)r(t)$$

There are two requirements for the model inverse to be applicable for control:

1. The inverse of the system must exist and must be causal. This is accomplished (as outlined in Section 5.4.2) by having no delays in the transfer function representation of the model $n_k = 0$, alternatively, in the state-space representation, having non-zero feed-forward matrix in $y(t) = Cx(t) + Du(t)$, $D \neq 0$. Logically, the inverse of a system with a delay is non-causal. Consider that at a time $t = t_k$ and a series of output measurements $y(t), t \leq t_k$, to determine the input signal $u(t), t \leq t_k$ that produced the output measurement would require information of the output in the future $y(t_{k+l})$, $l > 0$. Due to non-causality, the inverse of a system with a non-zero delay $n_k \neq 0$ does not exist.

2. The inverse of the system must be stable. This is accomplished in a discrete time system by having no transmission zeros outside the unit disk. The transfer function of the inverse of the system has poles at the original system's zeros, and zeros at the system's poles. Therefore, a system with unstable zeros has an unstable inverse.

The 14-parameter Output Error model (5.2) abides by the first requirement by construction, as the model parametrization was constrained to have $n_k = 0$. However, the transfer function has unstable transmission zeros, resulting in an unstable inverse. As such, the model is not appropriate for open-loop control. An alternative model is considered: using a 4-parameter model structure $n_b = 2$, $n_f = 2$, and $n_k = 0$, an optimized parametrization is found that has

neither unstable poles nor unstable zeros.

$$y(t) = \frac{B(z)}{F(z)}u(t) + e(t)$$

$$B(q) = 0.02084 + 0.0147q^{-1}$$

$$F(q) = 1 - 1.861q^{-}1 + 0.8802q^{-}2$$

The simulation and prediction performance of this low-order model is poorer than that of the high-order 14-parameter model; the step response shows the settling time of the simulation does not quite match the measurement (simulating the model using either filtered or unfiltered input data, and comparing to the respective output data), however the overshoot magnitude does fit well, and the steady-state gain is correct. The transfer function frequency response matches only at very low frequencies.

Despite that some of the performance metrics of the 4-parameter model appear worse, the prediction error performs comparably to the high-order model. The prediction error 2-norm, calculated via Parseval's theorem as the integral under the prediction error spectrum, shows only marginally worse performance than the 14-parameter Output Error model.

Table 6.1. Table of performance results of the 14-parameter and 4-parameter models

Model order	$n_b = n_f = 2, n_k = 0$	$n_b = n_f = 7, n_k = 0$
Prediction error 2-norm	0.17	0.13
# of unstable zeros	0	3

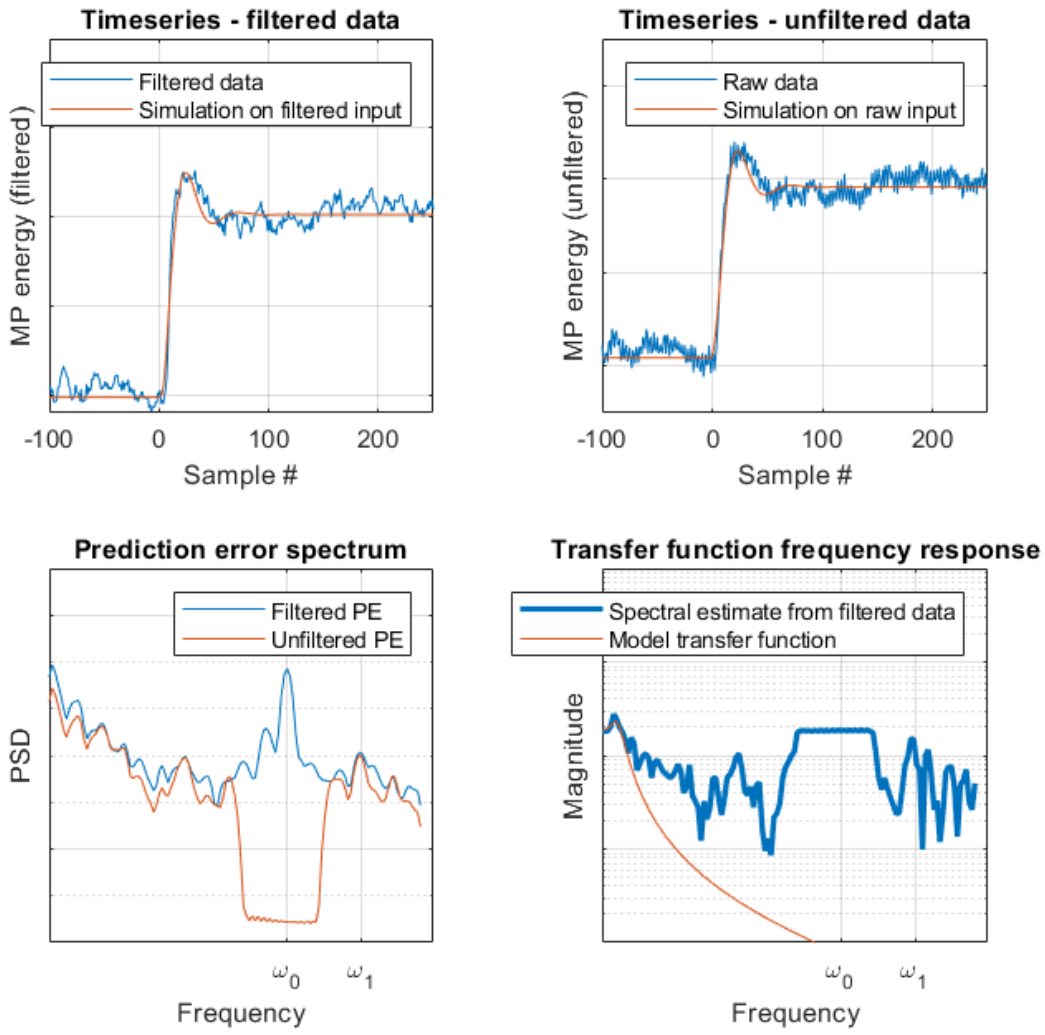


Figure 6.3. The model performance results of the 4-parameter model. Despite a lower-quality fit to the time series overshoot, the prediction error spectrum is comparable to that of the 14-parameter model. The 4-parameter model transfer function also does not amplify the 24kHz region, as seen in the 14-parameter model

This model, being that the transmission zeros are strictly stable, is invertible, therefore it can be used for open-loop control. The reuslts of the open-loop control demonstration

$$y(t) = G_\theta(q)u(t) = G_\theta(q)C(q)r(t) = G_\theta(q)G_\theta(q)^{-1}Q(q)r(t) = Q(q)r(t)$$

are shown below, using the trivial case of $Q(q) = 1$.

The desired output trajectory of the simulation can be a sinusoid with low amplitude, shown in the reference signal $r(t)$. This represents the desired change in amplified MP energy, at the output of the amplifier chain. The model inversion yields the control signal $u(t)$, then fed into the plant model results in the output signal $y(t)$ which exactly reflects the input signal $r(t)$. This is shown to result in a successful tracking of the reference signal in the below simulation.

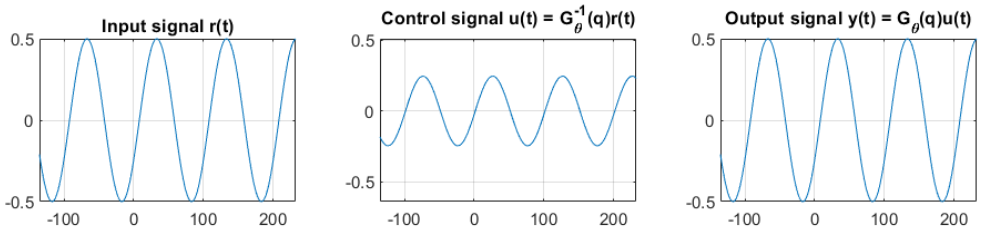


Figure 6.4. The open-loop control results, with the reference signal being a sine function. The output signal correctly tracks the input signal, using the intermediary control signal derived from the inverted plant model

While the model inversion is mathematically feasible, the control signal $u(t)$ that is necessary to achieve the step-function desired trajectory is impractically high. There are physical limits that constrain the range that the input Gain Command signal can reach. Extending the actuation range of the control signal input is not feasible beyond the current limits. This is shown to limit the feasible response that the amplifier can produce with a open-loop control setup.

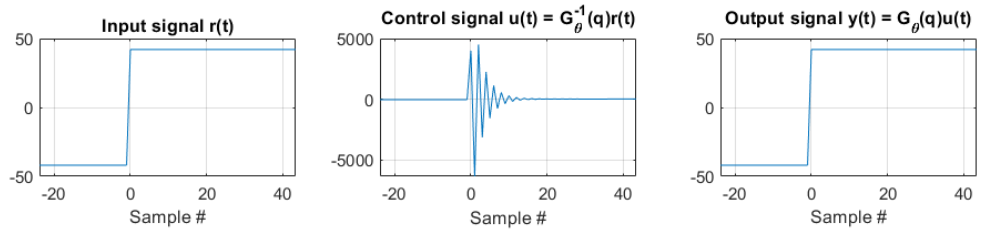


Figure 6.5. The open-loop control results, with the reference signal being a step function. The control signal necessary to result in a step function output substantially exceeds the practical bounds that the control input Gain Command can take

Chapter 7

Conclusions

The EUV light source drive laser is a complex machine, relying on precise controls and complicated physics to produce the laser beam that is critical to generating EUV-emitting plasma. The constituent elements of the drive laser, including the seed lasers, power amplifiers, timing control mechanisms, and metrology, are controlled with clockwork precision to generate a series of laser pulses with precisely the correct temporal placement, spatial positioning, and energy content in order to intercept the transient tin droplet and ionize it into an EUV plasma. The laser pulses, generated 50,000 times per second, are amplified to a 30kW power scale via the energized medium in the power amplifiers. The energy output of the amplifiers can be controlled via the Gain Command input. The energy of the amplified laser pulses is one of the most critical contributors to the power of the EUV light that the tin plasma generates, and the productivity of the EUV scanner relies on having the most intense, and most stable, EUV-emitting light source possible. As such, regulating the energy of the amplified laser pulses is critical for optimizing the EUV light source's performance.

In this work, the performance of the EUV light source drive laser has been investigated, to study the dynamic relationship between the Gain Command input and the amplified laser pulse energies. An experiment was executed on a prototype EUV light source to characterize the dynamic response of the laser energies to the input parameter, specifically investigating fast time-scale dynamics on the kilohertz scale. Several critical observations were made about the

laser performance, namely the second-order transient response to the Gain Command step input, and the high-frequency jitter mode present in the laser pulse energy spectrum.

The dynamic relationship between the Gain Command input and laser pulse energies was modeled using a parametric system identification framework. A discrete time linear time-invariant system description $y(t) = G(q)u(t) + H(q)e(t)$ was used as the overall structure of the model. The model transfer functions $G(q)$ and $H(q)$ are parametrized, with parameters optimized such that the prediction of the model matched the measured data as closely as possible, via minimizing the prediction error 2-norm. The model structure and model orders were selected such that the optimized parametrization produced the desired simulation results. In the identification performed on the experimental data from the high-power drive laser system, the second-order overshoot and settling along with the suppression of the high-frequency noise mode were used as the dictating factors to decide which model structure was considered best.

The high-frequency noise mode was suppressed from the data using a bandstop Chebyshev filter, which resulted in better optimization of the model parameters and improved model performance. The final optimized model was a 14-parameter Output Error model, which was shown to predict the measured laser energy response to the Gain Command input well.

Finally, the applicability of the model towards open-loop control was investigated. It was shown that the 14-parameter model, while demonstrating good prediction and simulation performance, was unstable when used in an open-loop controller. A 4-parameter model was found to yield similar prediction and simulation performance as the 14-parameter model, while maintaining stability in the open-loop controller configuration. This successfully showed that the system identification process as outlined in this work can be used to facilitate stable open-loop control of the system.

7.1 Next steps

7.1.1 Closed-loop control

With the simplified open-loop controller having been shown to be feasible, a closed-loop controller using the model inversion framework can be considered. In Hespanha (2018) [16], a closed-loop feedback controller, including the inverted plant model, is used to implement an arbitrary transfer function relationship between reference signal and output signal $y(t) = Q(q)r(t)$, which is robust against modeling errors between the physical plant model and the parametrized model. This closed-loop framework should be considered in the practical implementation of a controller, using the parameterized plant dynamics.

Ultimately, the reliance on system inversion in determining a control strategy poses challenges toward the utility of the controlled system. By inverting the plant model in the process of determining an controller, the frequencies where the plant model has low gain will translate into points where the inverse model has high gain. By prefiltering the data to remove known exogenous contributions, as done in the drive laser experiment data to remove the high-frequency component, the filtered plant model will deliberately have low gain at the filtered frequencies. Inversion will amplify these regions that we deliberately exclude.

Furthermore, a fundamental constraint on the implementation of a controller (whether open-loop or closed-loop) is the limited control amplitude. As shown in Figure 6.5, a desired trajectory of the system, given a linear system model, may require a control signal that exceeds the physical limit of the controller. A feedback control design which admits constraints, such as Model Predictive Control with full-state feedback, could be used in order to accommodate the limitations of the input signal.

7.1.2 Continuous-time modeling

The system identification process relied on using the Chebyshev filter to constrain the "unwanted" dynamics of the high-frequency jitter to the noise model, as this mode was found to

originate in the timing control system, and was unrelated to the Gain Command input response. The jitter could be incorporated into the plant model, however, via continuous-time modeling. A continuous-time model of the laser system can be re-derived from the discrete time transfer function representation of the model, as covered in Section 3.1.1. Whereas the current description of the discrete-time system assumes a constant inter-sample timing, the continuous-time model can be expanded to include the variable timing.

Separately, as the continuous-time dynamics are found to be periodic, a higher-order discrete time model could also be used to approximate these dynamics via a "lifting" approach. This process is covered in Farhood, Beck, and Dullerud [18].

7.1.3 Applications toward the EUV light source

The system identification methodology presented in this work can be used to develop new models of dynamics present in the EUV light source. Beyond this work, which investigated the relationship between the Gain Command parameter and the Main-Pulse energy, the methodology can also model the Gain Command to Pre-Pulse laser energy relationship. A more complicated model structure can be considered which incorporates more inputs and states of the drive laser, including the input energies of the seed lasers, as well as the timing control signals. Including more observables of the drive laser into the model will make the model more robust to varying EUV light source operating conditions. This framework could help prescribe new diagnostics to understand the EUV light source performance or new controllers that improve EUV stability, ultimately aiding in increasing the EUV scanner's productivity.

Furthermore, this research highlights the interstitial complexity of the drive laser amplification process. Beyond recommending new control apparatuses, improved laser energy regulation can be achieved by improving the intrinsic stability of the timing control of the laser generation process, which is synchronized to the tin droplet generation. Improved stability of the laser pulse energy generation is being developed for future EUV light source products.

Bibliography

[1] David C. Brandt, Igor Fomenkov, and Jayson Stewart, ”Progress in availability of NXE:3400B EUVL sources in the field and power scaling towards 500W (Conference Presentation)”, Proc. SPIE 11323, Extreme Ultraviolet (EUV) Lithography XI, 113230W (24 March 2020), https://doi.org/10.1117/12.2552424

[2] Peter Mayer, David C. Brandt, Igor Fomenkov, Michael Purvis, and Daniel Brown, ”Laser produced plasma EUV sources for N5 HVM and beyond: performance, availability and technology innovation”, Proc. SPIE 11609, Extreme Ultraviolet (EUV) Lithography XII, 1160918 (22 February 2021), https://doi.org/10.1117/12.2584407

[3] Karl Umstadter, Matthew Graham, Michael Purvis, Alex Schafgans, Jayson Stewart, Peter Mayer, Daniel Brown, ”EUV light source for high-NA and low-NA lithography,” Proc. SPIE 12494, Optical and EUV Nanolithography XXXVI, 124940Z (28 April 2023); doi: 10.1117/12.2657772

[4] Alexander A. Schafgans, Daniel J. Brown, Igor V. Fomenkov, Rick Sandstrom, Alex Ershov, Georgiy Vaschenko, Rob Rafac, Michael Purvis, Slava Rokitski, Yezheng Tao, Daniel J. Riggs, Wayne J. Dunstan, Matthew Graham, Nigel R. Farrar, David C. Brandt, Norbert Böwering, Alberto Pirati, Noreen Harned, Christian Wagner, Hans Meiling, Ron Kool, ”Performance optimization of MOPA pre-pulse LPP light source,” Proc. SPIE 9422, Extreme Ultraviolet (EUV) Lithography VI, 94220B (16 March 2015); doi: 10.1117/12.2087421

[5] Michael Purvis, Igor V. Fomenkov, Alexander A. Schafgans, Mike Vargas, Spencer Rich, Yezheng Tao, Slava I. Rokitski, Melchior Mulder, Erik Buurman, Michael Kats, Jayson Stewart, Andrew D. LaForge, Chirag Rajyaguru, Georgiy Vaschenko, Alex I. Ershov, Robert J. Rafac, Mathew Abraham, David C. Brandt, Daniel J. Brown, ”Industrialization of a robust EUV source for highvolume manufacturing and power scaling beyond 250W,” Proc. SPIE 10583, Extreme Ultraviolet (EUV) Lithography IX, 1058327 (10 May 2018); doi: 10.1117/12.2305955

[6] Levinson, H. J. (26 October 2020), Extreme Ultraviolet Lithography (Vol. PM326), SPIE Press. ISBN: 9781510639393; https://spie.org/Publications/Book/2581446

[7] Svelto, Orazio (1998). Principles of Lasers, 4th ed. (trans. David Hanna), Springer. ISBN 0-306-45748-2

[8] Siegman, A. E. (1986). Lasers, University Science Books. ISBN 0935702113

[9] Lee M. Frantz, John S. Nodvik; Theory of Pulse Propagation in a Laser Amplifier. Journal of Applied Physics 1 August 1963; 34 (8): 2346–2349. https://doi.org/10.1063/1.1702744

[10] Einstein, A. (1916). "Strahlungs-Emission und -Absorption nach der Quantentheorie" (trans. "Emission and Absorption of Radiation in Quantum Theory"). Verhandlungen der Deutschen Physikalischen Gesellschaft. 18: 318–323. Bibcode:1916DPhyG..18..318E. Translated in Alfred Engel. The Berlin Years: Writings, 1914-1917. Vol. 6. https://einsteinpapers.press.princeton.edu/vol6-trans/224

[11] Bewley, Thomas (2018). Numerical Renaissance: simulation, optimization, & control, Renaissance Press. ISBN # 978-0-9818359-0-7

[12] Ljung, Lennart (1987). System Identification: Theory for the User, Prentice Hall, Upper Saddle River, N.J. ISBN 10: 0138816409 / ISBN 13: 9780138816407

[13] Ljung, Lennart (2000). System Identification Toolbox for use with MATLAB. Version 5. The MathWorks, Inc, Nattick, MA, 5th Edition.

[14] de Callafon, Raymond A (Spring 2023). MAE283B Approximate Identification and Control, UCSD. Accessed August 21st, 2023. http://mechatronics.ucsd.edu/mae283b_23/index.html

[15] Ljung, Lennart (1999). Estimation Focus in System Identification: Prefiltering, Noise Models, and Prediction. Proceedings of the 38th Conference on Decision & Control

[16] Hespanha, João (2018). Linear Systems Theory (Second Edition). Princeton University Press, Princeton, New Jersey. ISBN-13: 978-0691179575

[17] Dautt-Silva, Alicia (2022). Robust Input Shaping for Linear Time-Invariant Models with Uncertainties. University of California, San Diego.

[18] Farhood, Beck, and Dullerud (2005). Model reduction of periodic systems: a lifting approach. Automatic 41 (2005) 1085-1090. doi:10.1016/j.automatica.2005.01.008