

Journal of Astronomical Telescopes, Instruments, and Systems

AstronomicalTelescopes.SPIEDigitalLibrary.org

Formulation of the production time and cost of the Lynx x-ray mirror assembly based on queuing theory

Jonathan W. Arenberg

SPIE.

Jonathan W. Arenberg, "Formulation of the production time and cost of the Lynx x-ray mirror assembly based on queuing theory," *J. Astron. Telesc. Instrum. Syst.* **5**(2), 021016 (2019), doi: 10.1117/1.JATIS.5.2.021016.

Formulation of the production time and cost of the Lynx x-ray mirror assembly based on queuing theory

Jonathan W. Arenberg*

Northrop Grumman Aerospace Systems, One Space Park Drive, Redondo Beach, California, United States

Abstract. The timely and predictable cost of the Lynx x-ray mirror assembly (XMA) is an essential element of the mission concept. We present an analytic model for the cost, schedule, and risk for the manufacture of a generalized system of many parts, and apply it to preliminary data of the manufacturing process for the XMA. The manufacturing process is modeled as a series of $G/G/w$ queues. The optimization of the manufacturing process, to minimize total process time, comes from the selection of the value of w , the number of servers performing each step in the manufacturing process, to avoid bottlenecks and minimizing idle servers. This analysis also includes the effects of finite process yield on cost and schedule. The cost model is parameterized by the various elements of cost, including the production time, thus linking the cost and schedule models. The system of coupled equations is the cost and schedule model. The process data that must be collected on the manufacturing process during the ongoing technology development process such as process times, yields, and distributions is identified. We conclude with the next steps that will be taken to make this analysis more complete.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JATIS.5.2.021016]

Keywords: cost and schedule analysis; cost optimization; Lynx; x-ray mirror assembly.

Paper 18125SS received Dec. 13, 2018; accepted for publication Apr. 12, 2019; published online May 14, 2019.

1 Introduction

Each of the four major decadal missions currently underway to support the Astro 2020 decadal review¹ has a major challenge to overcome for that mission to be determined to be executable. The existential problem is called the big fundamental problem (BFP) of the concept. The BFP for Lynx is the cost-effective manufacture of the Lynx x-ray mirror assembly (XMA) of suitable quality to meet the science objectives of the mission. The XMA is made of ~150,000 parts, 40,000 mirror elements, and 111,000 posts.² The technology to manufacture optical elements and posts is making excellent progress and is reported elsewhere in this volume.² The technical maturation is necessary but not sufficient to make Lynx viable. A strong case for the determinative cost and schedule for the revolutionary XMA is needed to provide sufficiency to the argument. This paper describes the analytic model and a few example of the analysis possible to demonstrate that such quantitative and determinative management is possible and executable.

All large hardware programs face the challenge of managing cost and schedule. In order to quantitatively manage any aspect of system performance, be it image quality, mass, power, thermal margin, or production schedule, an analytic model is needed. The technical performance metrics have models and tools to manage, but not schedule or cost. This paper will show that there is a solid analytic framework for establishing a rigorous management process to produce the XMA in a cost-efficient and effective manner. Failure to imagine this analytic model makes the analysis of strategic options and risks challenging at best and cost and schedule an exercise in reporting, not proactive management. The model to be developed in

this paper will be used during Lynx's development to analyze the schedule and cost for risk and allow for optimization later in Lynx's maturation.

This paper presents the derivation of the system of equations that predict the cost and duration of manufacture of the XMA. Analysis of the resulting system of equations identifies which parameters must be measured and determined during the ongoing technology development efforts. The uncertainties in the various parameters and the model assumptions are identified as the risks to cost and schedule performance.

The foundation of the modeling is an area of operations research known as queuing theory (QT). Operations research began in World War II and is described as "math and physics in service to the corner office."³ This is certainly such an application. Specifically, QT is the study of lines or queues that are an integral part of modern life.⁴ Queue or QT has been applied to health services,⁵ telephone network design, and many other important and interesting applications.⁶

The development of the approach begins with the introduction of useful notation and nomenclature, in which we will frame the development of the model. We will also introduce the definitions of the relevant terms for our discussion. The model development begins by building on some initial results presented in earlier work expanding the illustrative, but nonrealistic, example of an entirely deterministic XMA manufacturing process. The assumption that the production line is governed by a deterministic processes is then relaxed and the effects of variance are introduced, generalizing the model. The model for the cost and duration of an optimized process is examined in light of the Lynx technology development. This paper concludes with lessons that can be gleaned from the model even at the early stage of development and quantification. Especially important will be the identifications of the key assumptions and values that exert the greatest influence on cost, duration, and risk.

*Address all correspondence to Jonathan W. Arenberg, E-mail: jon.arenberg@ngc.com

The modeling and analysis of cost and schedule is not entirely new to the x-ray community. The process used to develop and optimize the AXAF (Chandra) x-ray calibration test schedule and efficiency⁷ is very similar to the approach discussed in this paper.

2 Nomenclature and Notation

The model being developed is based on QT, which is a well-developed area of applied mathematics and field of operations research. This section is intended to serve as a primer to introduce the QT nomenclature and notation that is used elsewhere in this paper.

The basic building block of our model development and analysis is the concept of a queue, a frequent occurrence in modern life. Such examples are airport security, fast food restaurants, rental car counters, and the infamous call centers for help with purchased items and services. Consider Fig. 1, a queue/server system, which will be shortened to being called a queue. Figure 1 shows that incoming elements, from the $i - 1$ th process, are processed in i 'th server, and once processed, they move on to the next stage of the process, the $i + 1$ th step. If the server is not busy, the element or part is processed and moves on to the next step. If the part from step $i - 1$ arrives and the server is busy, the incoming part waits in the queue or line for its turn. The mean population of the queue, the mean number of parts waiting for service, denoted L_q , is a key parameter in this analysis. The components of this simple system are the elements or parts, the queue where waiting is done and the server or service node or machine, which processes parts for the i 'th step and sends them on to the next step.

Figure 2 introduces some additional components to the system, namely the arrival pattern or distribution in time of the arriving element, denoted A , the distribution of service or process times, which is denoted B and the number of servers w . Kendall⁸ introduced a useful and compact notation to describe queues. In the eponymous notation, the queue is written $A/B/w$. (In more advanced analysis, the Kendall notation contains more parameters, but for present purposes the three attributes are sufficient.) The dummy variables A and B will be replaced with symbols representing various options of the statistical properties of the arrival and service processes. As we will see, the optimization of the time to manufacture is the story of how the w is chosen for each of the process steps.

The model of the production process is depicted in Fig. 3. This model is a series of queues that are serial, as a manufacturing process has a specific order, which cannot be altered. In this model, there are n steps and n shipments or transitions from one process to the next.

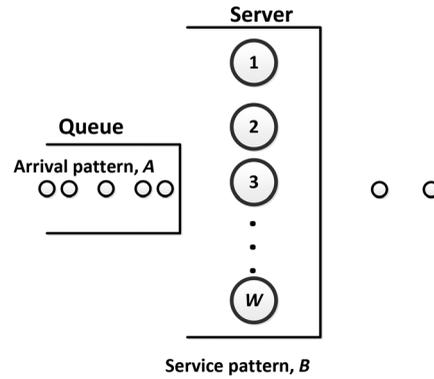


Fig. 2 Kendall's notation, A encodes the arrival distribution, B encodes the service time distribution, and w is the number of servers.

3 Deterministic Arrival and Service Times

The first model we will develop is the most naive model of an n -step process. Namely, the arrival model and the processing time models are deterministic (no variance) and denoted as D . In this case, a deterministic model means a specific time for each operation. We can now write compactly and efficiently that our process is made up of $n D/D/w_i$ queues, where the index i representing the ordinal value of each step.

3.1 Total Process Time: Determinative Process

Consider the time for a part to complete the n step process, this is called the total processing time, or T_P and is given

$$T_P = t_1 + s_1 + t_2 + s_2 + \dots + t_n + s_n, \tag{1}$$

t_i and s_i are the individual process and shipping times and $i = 1$ to n , indexing each step.

T_P is the processing time for a single element or unit, but Lynx will have to make many or Q units. To calculate the total process time for Q units T_Q , we need to know which process step has the largest process time. Define T_G as

$$T_G = \max(t_i, i \in [1, n]). \tag{2}$$

If a part is sent to the manufacturing process as soon as the first server is available, at some point in the process, the stream of parts reaches the step with the longest process time, the so-called gate, or rate limiting step or process. When the parts arrive at this gate, they arrive faster than they can be processed, and a queue builds up until the Q 'th part has arrived. Using Eqs. (1) and (2), an expression for T_Q can be written down as

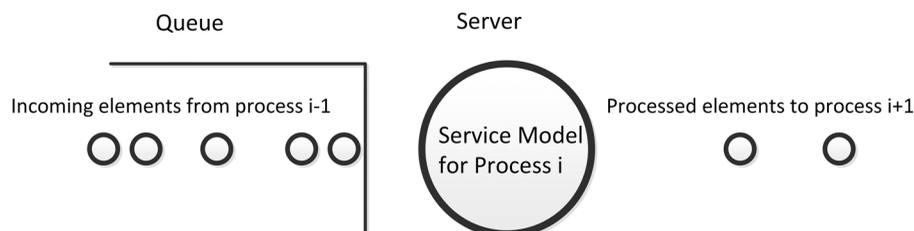


Fig. 1 Basic elements of a queue.

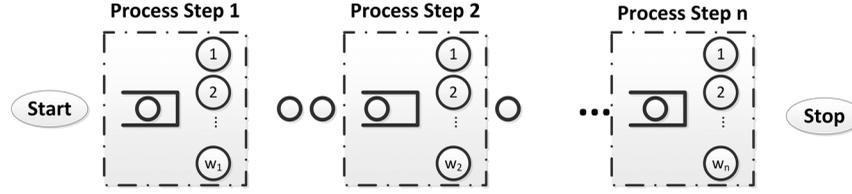


Fig. 3 Production line as a series of $A/B/w$ queues.

$$T_Q = \sum_{i=1}^{G-1} t_i + s_i + QT_G + s_G + \sum_{i=G+1}^n t_i + s_i, \quad (3)$$

where in the process, the gate G occurs does not matter, addition is associative so Eq. (3) is generally correct. Equation (3) can be rearranged to give

$$T_Q = \sum_{i \neq G} t_i + \sum_{i=1}^n s_i + QT_G. \quad (4)$$

In light of the formulation of Eq. (4), the name “gate” for the maximum process time becomes clear, it is the one process time that is multiplied by the number of parts to be made, dominating T_Q . If we wish to reduce T_Q , Eq. (4) shows us that the best investment is increase the number of servers that perform the process at the gate, effectively reducing the T_G . If the number of servers is w , then T_Q is given by

$$T_Q = \sum_{i \neq G} t_i + \sum_{i=1}^n s_i + \frac{QT_G}{w}. \quad (5)$$

Equation (5) is a good answer if the reduced gate time is much larger than all the others, namely

$$\frac{T_G}{w} \gg \max(t_i; i \neq G). \quad (6)$$

A simple and powerful lesson to be gleaned from this very naïve analysis is that knowledge of the values of the t_i allow for the most impactful investment. For example, if a second copy of the gate server is bought, the effective gate time is cut in half for the cost of that server. If we did not know this and bought a second complete copy of the production line, we would get essentially the same result, a reduction in T_Q by a factor of two, but at a cost of all of the equipment, a far less cost-effective strategy!

In order to formulate a more general solution, namely when Eq. (6) is not true, a change in notation is helpful. The author hopes that the reason for the change in notation, while not immediately obvious will be clear to the reader very shortly. Recall Eq. (1), which is rewritten

$$T_P = t_1 + s_1 + t_2 + s_2 + \dots + t_n + s_n. \quad (7)$$

Relabel the t_i according to their magnitude, denote the maximum value as $t_{\{1\}}$, the second largest process time as $t_{\{2\}}$, and so on, with the smallest value being $t_{\{n\}}$. The use of the curly braces is used so a reader is not confused with rank statistics traditional notation, which uses parentheses around the index and the index 1 indicates the minimum. The use of the braces hopefully will remind the reader that this is a reverse rank index. The assumption that all of the shipping times are small implies that all the XMA manufacturing will be done

at a single location. This co-location enables rapid transit from one step to the next and avoids transportation as the rate limiting process. Let $\sum_i s_i = S$, then Eq. (7) can be written

$$T_P = \sum_{i=1}^n t_{\{i\}} + S. \quad (8)$$

Equation (5) can be rewritten as

$$T_Q = \frac{Qt_{\{1\}}}{w_{\{1\}}} + \sum_{i=2}^n t_{\{i\}} + S. \quad (9)$$

Recall the condition, Eq. (6), and now consider increasing the value of $w_{\{1\}}$ until $t_{\{2\}} > \frac{t_{\{1\}}}{w_{\{1\}}}$, now the second longest process time is the gating process and Eq. (9) is no longer correct and $t_{\{2\}}$ is the gate, so an investment should be made in buying more servers at that step, and we have to also invest in addition servers for the original gate or it might regain its place as the longest step. We need to find $w_{\{1\}}$ and $w_{\{2\}}$ such that the effective process time is smaller than $t_{\{3\}}$ namely

$$\frac{t_{\{1\}}}{w_{\{1\}}} = \frac{t_{\{2\}}}{w_{\{2\}}} = t_{\{3\}}. \quad (10)$$

Hopefully, the reader can see the pattern that is emerging. Additional copies of the servers for the mitigation of gates or bottlenecks should consider how many levels b of bottlenecks are to be mitigated by investment.

If there are b bottlenecks to be mitigated, then Eq. (10) can be generalized to

$$\frac{t_{\{1\}}}{w_{\{1\}}} = \frac{t_{\{2\}}}{w_{\{2\}}} = \dots = \frac{t_{\{b\}}}{w_{\{b\}}} \leq t_{\{b+1\}}. \quad (11)$$

The mitigation of b bottlenecks gives T_Q as

$$T_Q = \sum_{i=1}^b \frac{t_{\{i\}}}{w_{\{i\}}} + Qt_{\{b+1\}} + \sum_{i=b+1}^n t_{\{i\}} + S. \quad (12)$$

Equation (12) gives us the means to examine key programmatic questions, such as, “What is the investment needed to meet the schedule duration requirement of D ?” or “What investments must be made to decrease the production time?”

Equation (12) can be used to derive a conservative estimate of T_Q . Consider the limiting case of Eq. (12), which is the equality; all of the terms in the first summation in Eq. (12) can be replaced with $t_{\{b+1\}}$, which gives the inequality

$$T_Q \leq \sum_{i=1}^b t_{\{b+1\}} + Qt_{\{b+1\}} + \sum_{i=b+1}^n t_{\{i\}} + S. \quad (13)$$

Equation (13) simplifies to

$$T_Q \leq (Q + b)t_{\{b+1\}} + \sum_{i=b+1}^n t_{\{i\}} + S. \quad (14)$$

By definition

$$t_{\{b+1\}} \geq t_{\{i\}} \quad i \in (b + 2, n). \quad (15)$$

So we have

$$\sum_{i=b+2}^n t_{\{i\}} \leq [n - (b + 2)]t_{\{b+1\}}. \quad (16)$$

Substitution of Eq. (16) into Eq. (14) now gives

$$T_Q \leq (Q + b)t_{\{b+1\}} + [n - (b + 2)]t_{\{b+1\}} + S. \quad (17)$$

Collecting terms gives

$$T_Q \leq (Q + n - 1)t_{\{b+1\}} + S. \quad (18)$$

In order to meet the hypothetical requirement on manufacturing span of D , we desire $T_Q < D$, this will also be met if

$$(Q + n - 1)t_{\{b+1\}} + S < D. \quad (19)$$

Equation (19) can be solved to give

$$t_{\{b+1\}} < \frac{D - S}{(Q + n - 1)}. \quad (20)$$

Equation (20) provides a means for estimating the w_i as will be shown in Sec. 8 of this paper.

3.2 Cost of Servers: Deterministic Case

Let c_i be the cost of the server for the $\{i\}$ th step, so the total cost of servers, C_S is

$$C_S = \sum_{i=1}^b w_{\{i\}} c_i + \sum_{i=b+1}^n c_i. \quad (21)$$

Using Eq. (11), the $w_{\{i\}}$ can be expressed in terms of the $t_{\{i\}}$:

$$w_{\{i\}} = \left\lceil \frac{t_{\{i\}}}{t_{\{b+1\}}} \right\rceil, \quad (22)$$

where the square brackets ($\lceil x \rceil$) indicate that the value x is rounded up to the next larger integer, as machines come in integral quantities. So C_S can be written as

$$C_S = \sum_{i=1}^b \left\lceil \frac{t_{\{i\}}}{t_{\{b+1\}}} \right\rceil c_i + \sum_{i=b+1}^n c_i. \quad (23)$$

4 Statistically Distributed Arrival and Service Times

Up to this point in the discussion, all of the process times t_i or $t_{\{i\}}$ and the s_i have been considered as deterministic quantities.

This section extends the deterministic analysis to accommodate stochastic arrival and service times. We introduce the additional concepts and results and develop the calculation for general distributions of arrival and service times including the effect of variance. Considerations of problems of this class are in effect the “bread and butter” of QT, and we have a rich heritage of results from which to draw on.

The foundation of QT is known as Little’s rule⁹⁻¹¹ and is written

$$L = \lambda W, \quad (24)$$

$$L_q = \lambda W_q, \quad (25)$$

and

$$W = W_q + \frac{1}{\mu}, \quad (26)$$

where L is the mean number in the queue/server system, λ is the mean rate of arrivals, μ is the mean service rate, W is the mean waiting time, and the subscript q refers to population and wait time in the queue, awaiting service. If the value of $W_q = 0$, the results of the deterministic queues are recovered. A seminal new parameter is also introduced here, called the server utilization, ρ and is defined

$$\rho = \frac{\lambda}{\mu}. \quad (27)$$

Small values of ρ mean that the arrival rate, λ is small compared to the service rate, and the server is not really very busy. If the values of ρ are large, then the server is busy most or all of the time, and there is a wait for service, or what we have termed a gate or bottleneck.

For a single server, if the arrival model is described by a Poisson process and the service time is exponential, the queue is described as $M/M/1$, where M means Markovian. For this case, it can be shown in Ref. 12 that L_q is given by

$$L_q = \frac{\rho^2}{1 - \rho}. \quad (28)$$

Result [Eq. (28)] states that to keep queue populations small, no bottlenecks, that ρ should be small. Figure 4 shows a plot of

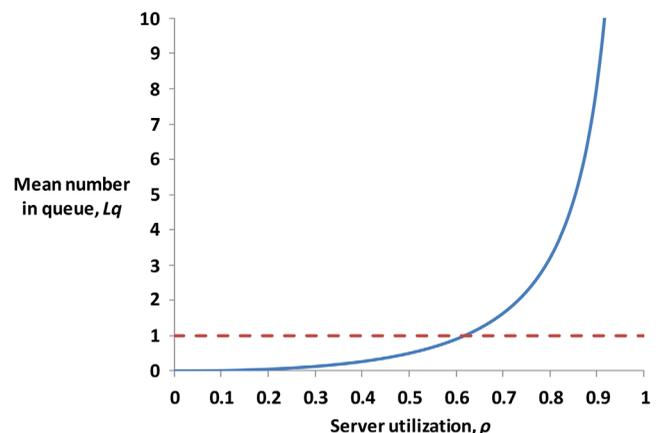


Fig. 4 L_q for an $M/M/1$ queue, Eq. (28).

Eq. (28) and shows that as ρ increases L_q diverges as $\rho \rightarrow 1$. Our strategy to minimize T_Q has been to avoid bottlenecks, in which parts are idle, now are described by the continuous variable ρ .

The statistical model of service times is not general, so consider now a $M/G/1$ queue, where the G meaning a general statistical distribution model of service times, characterized by its mean μ , and standard deviation in service time σ_s . L_q for this queue/server system, $M/G/1$, is known as the Pollaczek–Khintine (PK) formula^{13,14} and is given by

$$L_q = \frac{\lambda^2 \sigma_s^2 + \rho^2}{2(1 - \rho)}. \quad (29)$$

The PK result, Eq. (29), explicitly shows that service time variance does increase the mean number that are in the queue. Moreover, Eq. (29) informs us that we must measure and control σ_s and avoid large values.

The QT literature also provides a result for L_q in the case where both arrivals and service times are generally distributed.¹⁵ L_q is described by the closed form approximation

$$L_q \cong \left(\frac{\rho^2}{1 - \rho} \right) \left(\frac{1 + C_s^2}{2} \right) \left(\frac{C_a^2 + C_s^2}{1 + \rho^2 C_s^2} \right), \quad (30)$$

where C_x^2 means the coefficient of variance of x , which is the ratio of the variance to the mean squared. So C_s^2 and C_a^2 the coefficients of variance in service time, subscript s , and the variance in arrival time, subscript a , are given, respectively, by

$$C_s^2 = \frac{\sigma_s^2}{\left(\frac{1}{\mu}\right)^2} = \sigma_s^2 \mu^2 \quad (31)$$

and

$$C_a^2 = \frac{\sigma_a^2}{\left(\frac{1}{\lambda}\right)^2} = \sigma_a^2 \lambda^2. \quad (32)$$

Substitution of Eqs. (31) and (32) into Eq. (30) gives

$$L_q \cong \left(\frac{\rho^2}{1 - \rho} \right) \left(\frac{1 + \sigma_s^2 \mu^2}{2} \right) \left(\frac{\sigma_a^2 \lambda^2 + \sigma_s^2 \mu^2}{1 + \rho^2 \sigma_s^2 \mu^2} \right). \quad (33)$$

We want to operate with $\rho < 0.4$, using this and the definition of ρ , Eqs. (27) and (33) can be rewritten as

$$L_q \cong \left(\frac{\rho^2 \mu^2}{1 - \rho} \right) \left[\frac{0.16(\sigma_a^2 + \sigma_a^2 \sigma_s^2) + \sigma_s^2 + \sigma_s^4}{2(1 + \rho^2 \sigma_s^2 \mu^2)} \right]. \quad (34)$$

Examination of Eq. (34) illustrates the key role of variance in the service time, in increasing L_q and pushing the process to a bottleneck. Equation (34) also shows that variance in the arrival rate does matter but is of secondary importance. The implication of Eq. (34) is quite clear. Variance in service and arrival times can and should be planned for in any manufacturing schedule. Underestimation of these quantities will cause schedule delays and cost increases. In other words, risk is not that the manufacturing process for Lynx, or any other project, will have variance in service times. The risk lies in underestimation of that quantity. Proper measurement and estimation of the distribution of variance in service time is clearly one of the lessons from this

analysis that needs to be passed to the technology and manufacturing development efforts.

The result for L_q for an $M/M/w$ queue¹⁶ is found to be

$$L_q = \frac{P_0(\rho)^w \rho}{w!(1 - \rho)^2}, \quad (35)$$

where P_0 is given

$$P(\rho)_0 = \frac{1}{\sum_{m=0}^{w-1} \frac{(w\rho)^m}{m!} + \frac{(w\rho)^w}{w!(1-\rho)}}. \quad (36)$$

The result for W_q for a $G/G/w$ queue is given as¹⁷

$$W_q^{G/G/w} \cong W_q^{M/M/w} \left(\frac{C_a^2 + C_s^2}{2} \right). \quad (37)$$

Using Little's rule, Eqs. (25) and (37) give the expected queue population L_q for a $G/G/w$ queue:

$$L_q = \lambda \frac{P_0(\rho)^w \rho}{w!(1 - \rho)^2} \left(\frac{C_a^2 + C_s^2}{2} \right). \quad (38)$$

However, complicated a numerical process, Eq. (38), gives us the means of making a lower bound estimate for the number of servers w_i at each of the b bottlenecks chosen for mitigation. Namely, we seek the value of w_i that gives $L_q \ll 1$ given the other parameter values. Moreover, it can be seen that more servers do help and are a means to mitigate risk in the knowledge of the $C_a^2 + C_s^2$.

We are now in a position to calculate T_Q for the case of a production line made up of $G/G/w$ queues, with b bottlenecks mitigated. Recall the result for an $M/M/w$ queue, Eq. (18), mitigated to the b 'th level:

$$T_Q \leq (Q + n - 1)t_{\{b+1\}} + S. \quad (39)$$

To derive the result for a production line made up of n $G/G/w$ queues mitigated to the b 'th level, the gating time $t_{\{b+1\}}$ will be replaced with the total waiting time for the gating process, which includes the queue delay and call it Γ . Using Little's rule, Eq. (26) we get

$$\Gamma = W_q + \frac{1}{\mu_{\{b+1\}}}. \quad (40)$$

W_q is given by Eq. (25) and using Eq. (38) for L_q gives

$$\begin{aligned} W_q &= \frac{1}{\lambda} \lambda \frac{P_0(\rho)^w \rho}{w!(1 - \rho)^2} \left(\frac{C_a^2 + C_s^2}{2} \right) \\ &= \frac{P_0(\rho)^w \rho}{w!(1 - \rho)^2} \left(\frac{C_a^2 + C_s^2}{2} \right). \end{aligned} \quad (41)$$

So the expression for Γ is now and Eq. (40) becomes

$$\Gamma = \frac{1}{\mu_{\{b+1\}}} + \frac{P_0(\rho)^w \rho}{w!(1 - \rho)^2} \left(\frac{C_a^2 + C_s^2}{2} \right). \quad (42)$$

Equation (42) can be rewritten, if we recognize that the mean processing rate, $\mu_{\{b+1\}}$ is the reciprocal of the mean processing time, allowing us to write

$$\Gamma = t_{\{b+1\}} + \frac{P_0(\rho)^w \rho}{w!(1-\rho)^2} \left(\frac{C_a^2 + C_s^2}{2} \right). \quad (43)$$

Substitution of Eq. (42) into Eq. (39) gives the final result for TQ for G/G/w queues, mitigated to the b'th level as

$$T_Q \leq (Q + n - 1)\Gamma + S. \quad (44)$$

Using the equality gives the lowest upper bound, namely,

$$T_Q = (Q + n - 1)\Gamma + S. \quad (45)$$

Substitution of Eq. (43) into Eq. (45) gives the result for T_Q as

$$T_Q = (Q + n - 1) \left[t_G + \frac{P_0(\rho)^w \rho}{w!(1-\rho)^2} \left(\frac{C_a^2 + C_s^2}{2} \right) \right] + S. \quad (46)$$

Equation (46) shows that like in the deterministic case, the central role of the value of the gating process time is shown as well as the effects of variance on arrival and service times.

5 Cost Model

The analysis of schedule gives us the optimal number of servers and knowing the cost of each. It is a triviality to add that up to give the server cost, but that is not the whole cost of LMA production. The other elements that must be included are:

- LOE is the level of effort costs such as management, facilities rent, and utilities.
- I is the cost of installation of the servers and other factory modifications, clean room, electrical service, etc.
- U is the cost of operation of the factory, utilities.
- M is the cost of equipment maintenance.
- R is the cost of the raw materials.
- P is the cost of personnel working (not management).

The cost of XMA production is of course the sum of all these cost elements, namely

$$C_{XMA} = C_S + R + I + U + M + P + LOE. \quad (47)$$

We can now estimate how each cost element changes with respect to number of servers $W = \sum_{i=1}^n w_i$ and the time to produce the XMA T_Q. It is assumed that the cost of raw materials R is fixed with respect to production strategy. C_S is the cost of servers is from Eq. (21):

$$C_S = \sum_{i=1}^b w_i c_i + \sum_{i=b+1}^n c_i. \quad (48)$$

If we assume an average cost, to install servers, I is proportional to the number of servers W letting the average cost of installation \bar{I} , we get

$$I = \bar{I}W. \quad (49)$$

The cost of utilities U depends on the need from each server, times the number of server, and times the duration of the

production effort. If we assume the average cost of utilities for a server is u, then U is

$$U = uWT_Q. \quad (50)$$

Assuming the mean cost rate to maintain a server is m, then M is

$$M = mWT_Q. \quad (51)$$

Personnel costs are also proportional to the number of servers and the time to manufacture, so if the cost rate per server is p, then P is

$$P = pWT_Q. \quad (52)$$

The level of effort cost is proportion to the duration, if we call the LOW cost rate of l, then

$$LOE = lT_Q. \quad (53)$$

Substitution of Eqs. (48)–(53) into Eq. (47) gives

$$C_{XMA} = \sum_{i=1}^b w_i c_i + \sum_{i=b+1}^n c_i + R + \bar{I}W + uWT_Q + mWT_Q + pWT_Q + lT_Q. \quad (54)$$

Equation (54) can be rearranged and factored to give

$$C_{XMA} = \left(\sum_{i=1}^b w_i c_i + \sum_{i=b+1}^n c_i + R \right) + \bar{I}W + (u + m + p)WT_Q + lT_Q. \quad (55)$$

For any given set of w_i, the term in the parentheses in Eq. (55) is a fixed value, so Eq. (55) can be written

$$C_{XMA} = \alpha + \beta W + (\gamma + \delta W)T_Q. \quad (56)$$

We have seen above generally that if W is larger, there are more servers, but then T_Q will decrease, this means that it is possible to find a minimum value for the cost to manufacture the XMA C_{LMA}. The actual numerical calculation will have to be done once all the values of then relevant parameters have been determined.

6 Effects of Process Yield

Any real process is not perfect, and therefore has finite yield, where the yield for each step i is the probability that process i produces and acceptable result. The yield may be thought of as a probability of success for each process i and of course has the value from 0 to 1. Note that a part that fails to complete step i must be remade. This rework costs money in new material and time, causing T_Q to increase, which as we have seen in Eq. (56) also increases cost.

The yield for step i, y_i, is the probability of success, of step i, the probability of failure is the complement or 1 – y_i and since each process must product Q successes so the mean number of failures for each step i, f_i, is

$$f_i = Q(1 - y_i). \quad (57)$$

The total number of parts needing rework is the sum of Eq. (57) over all process steps i

$$F = \sum_{i=1}^n f_i = \sum_{i=1}^n Q(1 - y_i). \quad (58)$$

The probability of success or failure of a process is described by the binomial distribution, with Q large and $1 - y_i$ small, means small a rare set of failures so the special case of the Poisson distribution applies. The standard deviation of F , σ_F , is

$$\sigma_F = \sqrt{\sum_{i=1}^n Q(1 - y_i)}. \quad (59)$$

Using Eqs. (58) and (59), the number of additional parts that need to made to achieve Q successes with the associated risk that still more will be needed can be calculated. To achieve the required confidence, in general, more parts than F alone will need to be made. For the purposes of this discussion, assume to get the desired level of confidence we must make $F + v\sigma_F$ parts, which of course can be written as $F + v\sqrt{F}$. So the total time needed to make the LMA is the time to make $Q + F + v\sqrt{F}$ parts. Using Eq. (46), the total process time is

$$T_{\text{XMA}} = \left(Q + F + v\sqrt{F} + n - 1 \right) \left[t_G + \frac{P_0(\rho)^w \rho}{w!(1-\rho)^2} \left(\frac{C_a^2 + C_s^2}{2} \right) \right] + S, \quad (60)$$

which is also

$$T_{\text{XMA}} = \left\{ Q \left[1 + \sum_{i=1}^n (1 - y_i) + v \sqrt{\frac{\sum_{i=1}^n (1 - y_i)}{Q}} \right] + n - 1 \right\} \left[t_G + \frac{P_0(\rho)^w \rho}{w!(1-\rho)^2} \left(\frac{C_a^2 + C_s^2}{2} \right) \right] + S. \quad (61)$$

Equation (60) or Eq. (61) gives a formulation for the time to manufacture the LMA, the cost model, Eq. (56) when T_Q is replaced by T_{XMA} gives

$$C_{\text{XMA}} = \alpha + \beta W + (\gamma + \delta W) T_{\text{XMA}}. \quad (62)$$

The system of Eq. (60) or Eqs. (61) and (62) is the system that gives the cost and schedule estimates for the manufacture of the LMA.

7 Application Example: Determination of w_i

This section gives a simple example of how the results of this analysis can be applied to the problem of manufacturing the XMA. In discussion with the GSFC Optics Team,¹⁸ very early estimates of the process time for each step have been made. The results are presented in a table of t_i as Fig. 5.

Since this is very early in the development of the XMA manufacturing process, we will use results from Sec. 3, the determinative manufacture process, as the distribution of process times has not been characterized. To get an estimate of the w_i , the number of copies of servers for each step i , we start with Eq. (19) which gives the inequality

	Step	Name	Calendar time [h]
Element fabrication	1	CNC Grinding	4
	2	Lapping (buffing & measurement on FizCam)	8
	3	Slicing	3
	4	Coarse edge treatment	1
	5	Etch	1
	6	Polishing (FizCam)	20
	7	Smoothing (Fiz Cam)	4
	8	CNC Grinding	4
	9	Trimming	2
	10	Final Etching	3
			FizCam measurement, IBF, Fiz Cam
Coating	11	Measurement	3
	12	Cleaning, and oxidizing the backside	10
	13	Sputter Ir coating	10
	14	Anneal	10
	15	Fiz Cam Measure, IBF, Fiz Cam	3
Align & Bond	16	Measure height and locations of 8 spacers	0.5
	17	Fabricate 8 spacers	4
	18	Attach spacers and cure	2
		Measure radial heights of 8 spacers and trim to tolerance	
	19		3
	20	Align mirror by fine-grinding guided by Hartmann	4
		Bond mirror and cure, return to Step 16	
		complete module	10

Fig. 5 Table of preliminary XMA manufacturing steps and process times.

$$(Q + n - 1)t_{\{b\}} + S \leq D. \quad (63)$$

Equation (63) can be rewritten with the effective value of process time of the b 'th step as

$$(Q + n - 1) \frac{t_{\{b\}}}{w} + S \leq D, \quad (64)$$

as a means to estimate the value of w for mitigated step. We know from earlier in this paper, that all of the longer steps must have their w large enough that all $\frac{t_{\{i\}}}{w_i} \leq \frac{t_{\{b\}}}{w_b}$. This gives a simple means of evaluating the w and determining b . We select the limit case (equality) of Eq. (64) and solve for w_i giving

$$w_i = (Q + n - 1) \frac{t_i}{D - S}. \quad (65)$$

An interesting question can be posed using Eq. (65), "if we want to complete XMA production in less than D (to have planned margin), what are the w_i for those cases?" To address this question, D is replace by $D(1 - f)$, where is fractional reduction in time. Equation (65) now becomes

$$w_i = (Q + n - 1) \frac{t_i}{D - S}. \quad (66)$$

Let $D = 3$ years, further, we will assume that the shipping times are not the rate limiting steps so $S \sim 0$. From Fig. 5, we can see that $n = 21$ and we know that Q is 37,492.¹⁹ Figure 6 shows the evaluation of Eq. (66) for values of f from 0% to 50%.

For each value of $D(1 - f)$, the total number of servers has been determined and plotted against $D(1 - f)$ and shown in Fig. 7.

The trend line fit to the data in Fig. 7 is a quadratic. This result informs us that the sensitivity (partial derivative of duration with respect to cost) can be expected also to be quadratic as many costs are proportional to the number of servers. Such

Step #	t_p [hours]	w_i for $D(1-f)$					
		$f=0\%$	$f=10\%$	$f=20\%$	$f=30\%$	$f=40\%$	$f=50\%$
16	0.5	1	1	1	2	2	2
4	1	2	2	2	3	3	3
5	1	2	2	2	3	3	3
9	2	3	4	4	5	5	6
18	2	3	4	4	5	5	6
3	3	5	5	6	7	8	9
10	3	5	5	6	7	8	9
11	3	5	5	6	7	8	9
15	3	5	5	6	7	8	9
19	3	5	5	6	7	8	9
1	4	6	7	8	9	10	12
7	4	6	7	8	9	10	12
8	4	6	7	8	9	10	12
17	4	6	7	8	9	10	12
20	4	6	7	8	9	10	12
2	8	12	13	15	17	20	23
12	10	15	16	18	21	24	29
13	10	15	16	18	21	24	29
14	10	15	16	18	21	24	29
21	10	15	16	18	21	24	29
6	20	29	32	36	41	48	58

Fig. 6 Number of servers needed for each manufacturing step for varying values of f .

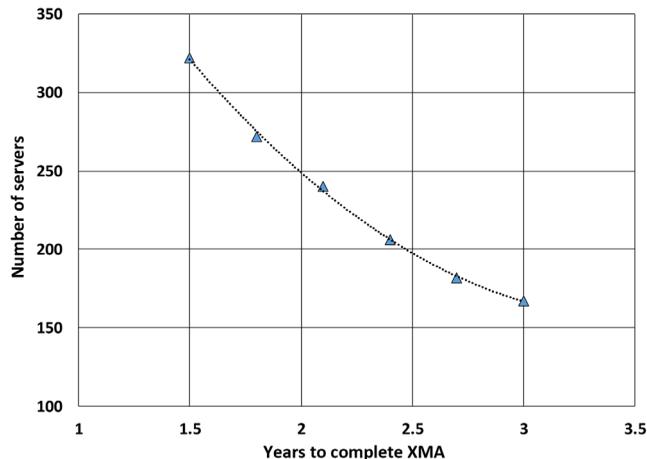


Fig. 7 Number of servers needed to time to complete XMA in varying times.

proportional costs are floor space, utilities, manufacturing personnel, and of course, the server and installation costs themselves. These investments decrease T_Q the time to manufacture the XMA. The optimal point can and will be determined once all the relevant parameter values are known.

8 Summary and Next Steps

The main objective of this work was to demonstrate that a credible model for the cost and schedule for production the Lynx mirror assembly exists. The proof of this is on offer in results, Eq. (60) or Eqs. (61) and (62).

During the development of the model, the derivation has identified relevant parameters and laid out the role of uncertainty or variance in these quantities as they affect cost and schedule. This analysis has shown that uncertainty in schedule and process time increases the time to manufacture the LMA and increases cost.

As the technology development proceeds, all of the relevant process times must be measured along with the distribution of

process times, so that we can properly go from t_i to $t_{\{i\}}$ that is rank ordering the processes correctly. All of this must be done in the process and technology development period.

This analysis is not complete. The next factor to be included in the analysis will be the availability of the servers, namely how often are they operating or and not in a state of repair or calibration or other off line status. Inclusion of finite availability will further refine the selection of the w_i or numbers of servers needed. Future work will also include the application of propagation of errors to develop a formal "error budget" for the cost and schedule models. The error budget approach will enable the further rigorous refinement of estimates of risk in cost and schedule.

Preliminary application of the model with early process data shows that the relationship between the number of servers and time to complete the XMA is quadratic, as shown in Fig. 7.

It is the happy conclusion of this work that it is possible to write down a model for cost and schedule and carry out a tolerance analysis on the value and assumption yielding a nuanced causal estimate of risk. This model gives us insight into how the XMA manufacture process works, so we can understand how the process will fail in terms of cost and schedule. Armed with this model and its descendants, the Lynx project will truly be able to manage the process of XMA manufacture and not just report on it.

Acknowledgments

The author would like to acknowledge the anonymous reviewers whose in-scope suggestions for this paper have helped make it better. Their out-of-scope suggestions will help guide future publications building on this work. The author gratefully acknowledges that much of this work was done under the Lynx Cooperative Agreement Notice contract funded by NASA and the matching Northrop Grumman internal funding.

References

1. M. P. Nagaraja et al., "2020 Decadal survey planning," <https://science.nasa.gov/astrophysics/2020-decadal-survey-planning> (5 December 2018).
2. W. W. Zhang et al., "High-resolution, lightweight, and low-cost x-ray optics for the Lynx observatory," *J. Astron. Telesc. Instrum. Syst.* **5**(2), 021012 (2019).
3. P. M. Morse and G. E. Kimball, "Methods of operations research," OEG Report No. 54, 1st edn., National Defense Research Committee, Washington, DC (1946).
4. J. F. Shortle et al., *Fundamentals of Queueing Theory*, Wiley Series in Probability and Statistics, 5th ed., John Wiley and Sons, Hoboken, New Jersey (2018).
5. L. Mayhew and D. Smith, "Using queuing theory to analyse completion times in accident and emergency departments in the light of the Government 4-hour target," Cass Business School (2006).
6. A. Nafees, "Analysis of the sales checkout operation in ICA supermarket using queuing simulation," http://www.statistics.du.se/essays/D07E_Nafees.pdf (5 December 2018).
7. J. W. Arenberg et al., "Lessons we learned designing and building the Chandra telescope," *Proc. SPIE* **9144**, 91440Q (2014)
8. D. G. Kendall, "Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain," *Ann. Math. Stat.* **24**, 338–354 (1953).
9. J. D. C. Little, "A proof for the queuing formula: $L = \lambda W$," *Oper. Res.* **9**(3), 383–387 (1961).
10. W. S. Jewell, "A simple proof of: $L = \lambda W$," *Oper. Res.* **15**(6), 1109–1116 (1967).
11. S. Eilon, "A simpler proof of $L = \lambda W$," *Oper. Res.* **17**(5), 915–917 (1969).

12. D. Gross et al., *Fundamentals of Queueing Theory*, 4th edn., John Wiley and Sons, Hoboken, New Jersey (2008).
13. F. Pollaczek, "Über eine Aufgabe der Wahrscheinlichkeitstheorie," *Math. Z.* **32**, 64–100 (1930).
14. A. Y. Khintchine, "Mathematical theory of a stationary queue," *Mat. Sb.* **39**(4), 73–84 (1932).
15. W. G. Marchal and C. M. Harris, "A modified Erlang approach to approximating GI/G/1 queues," *J. Appl. Probab.* **13**(1), 118–126 (1976).
16. L. Kleinrock, *Queueing Systems Volume 1: Theory*, pp. 101–103, 404, John Wiley & Sons, Inc., New York (1975).
17. J. F. C. Kingman, "The single server queue in heavy traffic," *Math. Proc. Cambridge Philos. Soc.* **57**(4), 902 (1961).
18. W. W. Zhang, Email to author (12 March 2019).
19. J. A. Gaskin et al., "Lynx x-ray observatory: an overview," *J. Astron. Telesc. Instrum. Syst.* (in press).

Jonathan W. Arenberg is the chief engineer for space science missions at Northrop Grumman Aerospace Systems. He has worked on the Chandra x-ray observatory and James Webb space telescope. He co-conceived and developed the Starshade for the direct imaging of extra-solar planets. He is the author of more than 180 conference presentations, papers, book chapters, and a recent SPIE press book. He holds a dozen European and U.S. Patents and is an SPIE fellow.