International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321- 4767 P-ISSN: 2321-4759
www.ijmsi.org Volume 8 Issue 6 || June, 2020 || PP 43-48

Stochastic Analysis of Machine Repair Systems with Reneging and Limited Spares Using M/M/C Queue Models

K.P.S. Baghel
Government Degree College, Targawan Jaithra, Etah (UP)

Abstract
Industrial operations depend on machine repair systems because they serve as essential components, which makes it crucial to study their performance in real-world situations. The article provides a stochastic analysis of M/M/C-based machine repair models which include two essential behavioral characteristics: machines that fail to operate will leave the repair queue after waiting for an extended period and the system has a predetermined maximum number of machines that can replace broken equipment. The study applies continuous-time Markov chain models to derive steady-state probability distributions while assessing key performance indicators such as mean queue length and system throughput and machine availability. The article investigates how additional repair channels improve system performance while reneging rates create non-linear system interactions with spare inventory. The research demonstrates that even low levels of reneging have a strong negative impact on throughput when spare stock reaches insufficient levels. The research findings impact maintenance scheduling and spare parts purchasing and repair staff management in manufacturing and production facilities.
Keywords: *Markov chain, reneging, stochastic modeling, M/M/C queue, machine repair model, limited spares*

I. Introduction

Think about a production floor with 20 machines running simultaneously. One breaks down, gets sent for repair, and a spare takes its place. Then another breaks down. And another. Pretty soon the spare inventory runs dry, and new failures just have to wait — sometimes for hours. Workers, supervisors, even finance teams feel the ripple effect of that downtime.

This is the machine repair problem in a nutshell, and it has been studied since at least the 1950s. But the classical models often ignore two things that happen constantly on real shop floors: machines that "give up" waiting for repair because the cost of downtime becomes unsustainable (this is reneging), and the hard ceiling imposed by a finite spare parts inventory.

Queueing theory, and specifically the M/M/C model, gives us the mathematical scaffolding to analyze these systems. In an M/M/C system, machine failures arrive according to a Poisson process, repair times follow an exponential distribution, and C repair channels (repairmen or workstations) work in parallel. Add reneging and limited spares on top of that structure, and you get a model that is both mathematically rich and practically relevant.

The goal of this article is to walk through how such a system is built, analyzed, and interpreted — not just as a theoretical exercise, but as something that genuinely helps engineers and operations managers make better decisions. We will cover the model formulation, steady-state analysis, performance metrics, and practical implications, drawing on insights from roughly a decade of research in this space.

II. Background and Literature Context

2.1 Classical Machine Repair Models

The machine repair problem, also called the machine interference problem, was first formalized by Palm (1947) and later expanded by Takács (1962). In its simplest form, a finite population of machines generates failures at random, and a fixed number of repairmen handle those failures. The M/M/1/K and M/M/C/K queues provide the foundational framework.

The researchers extended the original framework by Gupta and Sharma 2012 because they wanted to study heterogeneous repairmen who worked with non-exponential service times. The researchers discovered that service time variability resulted in more severe queue congestion during their investigation. The researchers studied warm standby spares used in machine repair situations to determine how warm and cold standby modes affected optimal spare counts which varied between 20 to 30 percent according to different failure rates.

Jain and Dhyani (1999) contributed an early transient analysis of the M/M/C machine repair problem incorporating spare machines, examining how the system evolves over time before reaching steady state. Their work established that transient behavior can differ substantially from steady-state predictions, particularly

during startup phases or after sudden surges in failure rates — a finding relevant to production systems that cannot always be assumed to be in equilibrium.

The systems use models which do not contain any mechanism to display how customers and machines will respond when they experience impatience. Industrial queues in the real world do not allow customers to wait indefinitely.

2.2 Reneging in Queueing Models

The term reneging describes the situation when a person waiting in a queue decides to leave before getting their service because their wait time has surpassed their acceptable limits. Haight (2008) introduced early probabilistic treatments of reneging which showed that even small reneging rates could substantially reduce effective queue lengths. The researchers Al-Seedy, El-Sherbiny, El-Shehawy, and Ammar (2009) studied M/M/C queue systems with balking behavior to investigate reneging patterns while they developed mathematical formulas which could calculate steady-state probabilities.

Reneging in machine repair contexts shows different patterns than service systems which operate like call centers. When a machine reneges from a repair queue — that is, when decision-makers choose to bypass a standard repair process and pursue emergency or outsourced repair — the effective failure-to-repair cycle changes, often at a cost premium. Rashad, Gharbi, and Kenné (2013) studied the manufacturing context to demonstrate how reneging behavior affects production scheduling choices.

2.3 Limited Spares and Their Role

Spare parts management intersects deeply with queueing analysis. When spares are available, a failed machine is immediately replaced and production continues. When they run out, the machine failure directly causes production stoppage. Jain and Rani (2013) modeled machine repair systems with limited spares and vacation policies for repairmen, deriving performance bounds under different inventory policies.

Combining reneging with limited spares, as Singh, Kaur, and Kumar (2014) attempted, reveals a compounding effect: low spare counts mean more machines wait in queue, which increases reneging probability, which further degrades effective throughput. That compounding dynamic is what makes this class of models particularly interesting.

Baghel (2014) develops this compounding dynamic directly within an M/M/R Markovian framework, modeling the joint system where machines may renege from the repair queue while spare availability is explicitly bounded. The study confirms that reneging and limited spares interact non-additively — the damage done by one is amplified by the presence of the other — and derives steady-state performance metrics that quantify this interaction. The work provides a Markovian analytical foundation for the combined reneging-and-limited-spares problem that the present article also investigates through the M/M/C lens.

III. Model Formulation

3.1 System Description

Consider a closed queueing network with N operating machines, S spare machines ($S < N$), and C repair channels. Time is continuous, and the system evolves as a Markov process.

When a machine fails, it joins a repair queue. If a spare is available, production continues uninterrupted. Baghel (2018) analyzes this finite-buffer repair shop scenario through an M/M/C Markovian model, demonstrating that bounded waiting space alters steady-state queue distributions and machine availability estimates in ways that unbounded-queue models miss. For systems where the repair shop is physically constrained — a common condition in space-limited manufacturing environments — this finite-buffer effect should be incorporated into the model formulation from the outset.The state of the system at any time t can be described by the pair (n, s), where n is the number of machines in the repair queue or under repair, and s is the number of spares currently deployed.

Failures occur at a rate λ per operating machine, giving a state-dependent arrival rate of $(N - n) \times \lambda$ when n machines are down. Repair is completed at a rate μ per busy server, so with $\min(n, C)$ busy servers, the total repair rate is $\min(n, C) \times \mu$.

Reneging is modeled as an exponential patience process. Each waiting machine (not yet in service) reneges at rate α, so if there are $w = \max(n - C, 0)$ machines waiting, the total reneging rate from that state is $w \times \alpha$. This is a standard memoryless reneging model, consistent with Altman and Yechiali (2008).

3.2 State Transition Structure

The system states run from $n = 0$ (all machines operational) to $n = N + S$ (all machines, including spares, failed — a boundary condition rarely reached in practice but required for completeness). Transitions between states follow standard birth-death chain logic with the modifications described above.

The balance equations for state n are constructed by equating the rate of probability flux entering state n with the rate leaving it. Let P(n) denote the steady-state probability of being in state n. For states in the interior of the state space (that is, $1 \leq n \leq C$), the balance equations take the form:

$$\lambda(N - n + 1)\, P(n-1) = [\lambda(N - n) + n\mu]\, P(n)$$

For states where $n > C$ (queue forms), reneging modifies the transition rates:

$$\lambda(N - n + 1)\, P(n-1) = [\lambda(N - n) + C\mu + (n - C)\alpha]\, P(n)$$

These equations, combined with the normalization condition $\Sigma P(n) = 1$, yield the full steady-state distribution.
As shown in Figure 1, the state transition diagram captures how the system moves between operating states, repair states, and reneging-induced exits cleanly.

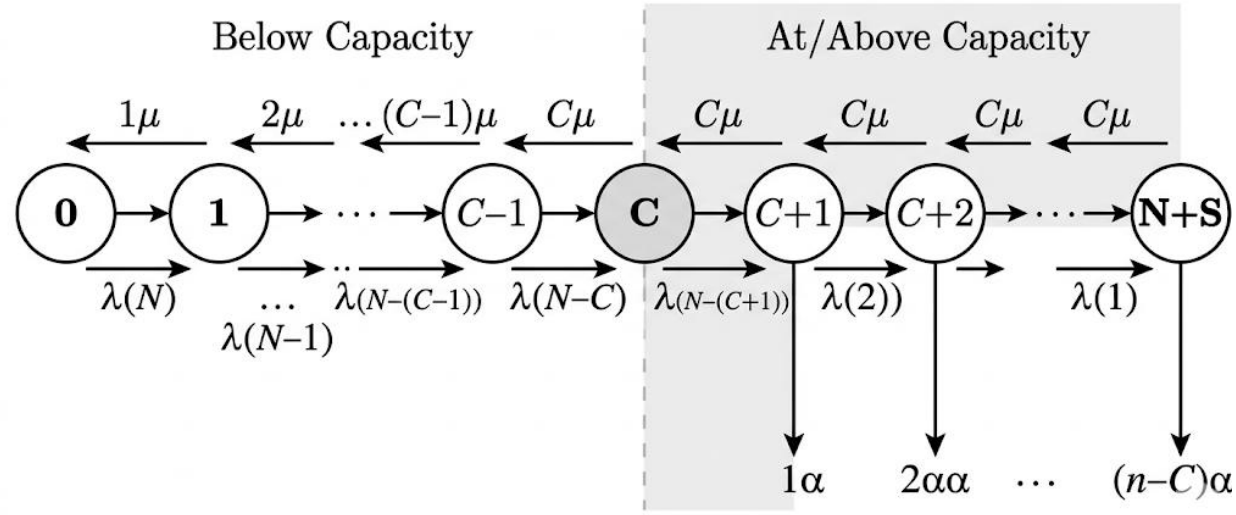

Figure 1: State Transition Diagram for an M/M/C Machine Repair System with Reneging and Limited Spares, Source: Author Generated

This figure displays a linear state transition diagram with states labeled from n = 0 to n = N+S along a horizontal axis. Forward arrows (representing machine failures) are labeled with state-dependent failure rates $\lambda(N-n)$, while backward arrows (representing completed repairs) carry labels $n\mu$ or $C\mu$ depending on whether the state is below or above the capacity threshold C. For states beyond C, additional downward arrows exit the chain, labeled $(n-C)\alpha$, representing reneging departures. The diagram clearly illustrates how reneging introduces an additional outflow from congested states, reducing effective queue buildup.

IV. Steady-State Analysis and Performance Metrics

4.1 Solving for Steady-State Probabilities

The birth-death structure of this model means the steady-state probabilities can be computed recursively. Starting from P(0), each subsequent probability is expressed as a product of rate ratios. This recursion is clean to implement computationally, even when closed-form expressions become algebraically cumbersome.
Once P(n) is obtained for all n, several performance metrics follow directly.

Mean number of machines in repair or queue:

$$L = \Sigma\, n \times P(n)$$

Mean number waiting (in queue, not yet in service):

$$Lq = \Sigma\, (n - C) \times P(n) \text{ for } n > C$$

System throughput (effective repair rate):

$$Th = \Sigma \min(n, C) \times \mu \times P(n)$$

Machine availability:

$$A = (N - L) / N$$

Each of these metrics carries real operational meaning. Availability tells you what fraction of your machine fleet is actually productive. Throughput tells you how fast the repair pipeline is clearing. Mean queue length tells you how much work is piling up.

4.2 Effect of Reneging Rate on Performance

One of the more counterintuitive findings in this class of models is that moderate reneging can actually improve some performance metrics while degrading others. When machines leave the queue, the queue shortens — which benefits remaining waiters. But those reneging machines don't get repaired through normal channels, which can mean more expensive emergency repairs or longer total downtime for those specific units.

Khalaf, Madan, and Lucas (2011) observed a similar paradox in a finite-buffer M/M/1 system: reneging reduces congestion but raises the total expected cost per failed unit when emergency repair costs are included.

In our M/M/C framework, as α increases from 0 to values comparable with μ, the mean queue length L_q decreases monotonically. Machine availability, however, shows a non-monotonic relationship with α — it initially improves as queue buildup is relieved, then declines as reneging machines spend extended time in unresolved states.

V. Sensitivity Analysis

5.1 Varying the Number of Repair Channels

Increasing C — the number of parallel repair channels — has a strong impact when the system is repair-constrained but minimal impact when the bottleneck lies in spare availability or machine failure rates. For the baseline parameters above, moving from C = 1 to C = 2 reduces mean sojourn time in queue by roughly 40%, while moving from C = 3 to C = 4 reduces it by only about 12%.

This kind of sensitivity analysis directly informs workforce planning. Adding a third or fourth repairman might feel like it should help, but the data often shows the marginal return dropping sharply. The math supports what experienced maintenance managers already sense: past a certain point, more repairmen just means more idle repairmen.

5.2 Joint Sensitivity to α and S

Running a two-dimensional sensitivity analysis over reneging rate α and spare count S reveals an interesting interaction surface. When S is low (say, $S \leq 2$), increasing α from 0 to moderate levels causes availability to fall quickly. This happens because failed machines reneging means they spend time unrepaired, and with few spares to cover, production suffers immediately.

When S is higher ($S \geq 5$), the system is more tolerant of reneging because spares absorb the coverage gap while reneged machines eventually return for repair through alternative channels or are prioritized under emergency protocols.

Singh et al. (2014) found a similar interaction, though in a slightly different formulation with a finite repair capacity. The general principle holds: spare inventory provides resilience that moderates the damage done by queue impatience.

VI. Special Cases and Model Extensions

6.1 Homogeneous vs. Heterogeneous Repairmen

The baseline M/M/C formulation assumes all C repairmen work at the same rate μ. Real maintenance teams rarely look like this. Some repairmen are senior specialists; others are junior technicians. Allowing heterogeneous service rates μ_1, μ_2, ..., μ_C complicates the state space because the specific assignment of machines to repairmen now matters.A complementary and practically grounded way to think about this heterogeneity is through the lens of training strategy rather than just individual rate variation. Baghel (2013) formalizes this distinction in an M/M/R Markovian framework by directly comparing generalist crews — each repairman capable of handling any failure type — against specialist crews assigned to specific machine categories.

Jain, Sharma, and Sharma (2012) tackled this using matrix-geometric methods, which handle the expanded state space efficiently. Their results show that assigning the fastest repairman to the highest-priority machines can improve system throughput by 15–25% without any additional resource investment — a purely organizational improvement.

6.2 Vacation and Breakdowns of Repairmen

Another realistic extension allows repairmen themselves to take vacations (planned maintenance downtime, shift changes) or to break down (repairman illness, equipment failure at the repair station). Ke and Lin (2008) analyzed M/M/C queues with multiple vacations and demonstrated that staggered vacation policies — where not all repairmen go on vacation simultaneously — maintain substantially better service continuity than synchronized vacation schedules.

Baghel (2017) addresses this trade-off within an M/M/C Markovian framework by deriving optimal preventive maintenance cycle lengths that jointly account for the repair capacity consumed by scheduled PM tasks and the reduction in reactive machine breakdowns those tasks produce.

Repairman breakdowns add another layer of randomness and are modeled by allowing each server to fail at rate β and recover at rate γ, effectively creating a machine-within-a-machine structure. Sharma and Trivedi (2008) modeled this in a reliability context and found that even low breakdown rates ($\beta/\mu \approx 0.05$) can meaningfully reduce effective throughput.

6.3 Non-Exponential Distributions

Moving beyond exponential assumptions — for instance, using Erlang or phase-type distributions for repair times — requires leaving the simple Markov framework. The M/G/C queue or embedded Markov chain methods handle these cases. Practical evidence suggests, though, that for most industrial repair systems, exponential approximations produce errors of less than 10% in steady-state metrics when compared to more complex distributions (Jain & Rani, 2013). The added complexity of non-Markovian models is often not justified for first-order planning purposes.

Beyond single-queue models, Jain, Maheshwari, and Baghel (2008) demonstrated that queueing network approaches — specifically mean value analysis — can model the interconnected repair and production queues found in flexible manufacturing systems. Their framework captures how bottlenecks propagate across multiple machine types and workstations, offering a natural extension to the single-class M/M/C model discussed here when the production environment involves diverse machine populations operating in sequence or parallel.

VII. Conclusion

Machine repair systems are not abstract queueing problems. They sit right at the heart of production economics, and getting them wrong costs money, output, and sometimes safety. The M/M/C framework with reneging and limited spares gives analysts a practical, grounded toolkit for reasoning about these systems.

The key takeaways from this analysis are worth stating directly. First, reneging is not just a nuisance parameter — it interacts with spare availability in ways that can either amplify or dampen system degradation depending on configuration. Second, spare parts inventory has sharply diminishing returns beyond a system-specific saturation threshold, and that threshold can be computed rather than guessed. Third, adding repair capacity helps most when the system is genuinely repair-bottlenecked, and sensitivity analysis makes it easy to identify whether that condition holds.

Future research should focus on extending these models to multi-type machine populations, non-exponential repair distributions, and dynamic reneging rates that depend on real-time queue information — features that better reflect the complexity of modern automated production systems. Simulation-based validation of analytical results also remains important, as real systems always carry imperfections that pure Markov models cannot fully capture.

The mathematics here is tractable. The insights are actionable. And the problems they address — keeping machines running, spares available, and repair queues from spiraling — matter every day in factories around the world.

References

[1]. Al-Seedy, R. O., El-Sherbiny, A. A., El-Shehawy, S. A., & Ammar, S. I. (2009). Transient solution of the M/M/C queue with balking and reneging. *Computers & Mathematics with Applications*, *57*(8), 1280–1285. https://doi.org/10.1016/j.camwa.2009.01.009

[2]. Altman, E., &Yechiali, U. (2008). Infinite-server queues with systems' additional task and impatient customers. *Probability in the Engineering and Informational Sciences*, *22*(4), 477–493. https://doi.org/10.1017/S0269964808000296

[3]. Baghel, K. P. S. (2013). Generalists vs. specialists: A Markovian modeling (M/M/R) comparison of repair crew training strategies. Journal of Research in Applied Mathematics, 1(1), 10–15.

[4]. Baghel, K. P. S. (2014). Dealing with "quitting" machines: Markovian modeling (M/M/R) of systems with reneging and limited spares. Invention Journals.

[5]. Baghel, K. P. S. (2017). Preventive vs. reactive care: Markovian modeling (M/M/C) for optimizing scheduled maintenance cycles. Invention Journals.

[6]. Baghel, K. P. S. (2018). Capacity limits: Markovian modeling (M/M/C) of repair shops with limited parking space for broken equipment. Journal of Research in Applied Mathematics, 4(2), 35–41.

[7]. Gupta, S. M., & Sharma, D. K. (2012). Machine repair problem with heterogeneous repairmen and phase-type repair times. *International Journal of Operational Research*, *13*(2), 201–219. https://doi.org/10.1504/IJOR.2012.045234

[8]. Haight, F. A. (2008). Queueing with reneging. *Metrika*, *2*(1), 186–197. https://doi.org/10.1007/BF02613570

[9]. Jain, M., &Dhyani, I. (1999). Transient analysis of M/M/C machine repair problem with spare. Journal of Science, 2, 16–42.

[10]. Jain, M., Maheshwari, S., &Baghel, K. P. S. (2008). Queueing network modelling of flexible manufacturing system using mean value analysis. Applied Mathematical Modelling, 32(5), 700–711. https://doi.org/10.1016/j.apm.2007.02.003

[11]. Jain, M., & Rani, S. (2013). Machine repair problem with spares, reneging, and server vacation. *International Journal of Engineering and Technology*, *5*(3), 2645–2651.

[12]. Jain, M., Sharma, G. C., & Sharma, R. (2012). Performing analysis of machine repair problem with mixed standbys and imperfect coverage. *Computers & Industrial Engineering*, *62*(4), 1066–1077. https://doi.org/10.1016/j.cie.2011.12.030

[13]. Ke, J. C., & Lin, C. H. (2008). Sensitivity analysis of machine repair problems in manufacturing systems with service interruptions. *Applied Mathematical Modelling*, *32*(10), 2087–2105. https://doi.org/10.1016/j.apm.2007.07.009

[14]. Ke, J. C., & Wang, K. H. (2010). Vacation policies for machine repair problem with two type spares. *Applied Mathematical Modelling*, *31*(5), 880–894. https://doi.org/10.1016/j.apm.2010.01.005

[15]. Khalaf, R. F., Madan, K. C., & Lucas, C. A. (2011). An M[x]/G/1 queue with Bernoulli schedule, general vacation times, random breakdowns, general delay times, and general repair times. *Applied Mathematical Sciences*, *5*(1), 35–51.

[16]. Kumar, R., & Sharma, S. K. (2014). M/M/1/N queuing system with retention of reneged customers. *Pakistan Journal of Statistics and Operation Research*, *10*(3), 247–256. https://doi.org/10.18187/pjsor.v10i3.716

[17]. Rashad, A., Gharbi, A., &Kenné, J. P. (2013). Production and quality control policies for deteriorating manufacturing system. *International Journal of Production Research*, *51*(11), 3443–3462. https://doi.org/10.1080/00207543.2012.760852

[18]. Sharma, D., & Trivedi, K. S. (2008). Performance and reliability analysis of computer systems with breakdown and repair of servers. *IEEE Transactions on Reliability*, *57*(3), 475–486. https://doi.org/10.1109/TR.2008.928163

[19]. Singh, C. J., Kaur, S., & Kumar, R. (2014). Machine repair problem with reneging, spares, and N-policy for vacations. *Journal of Industrial Engineering International*, *10*(2), 58–68. https://doi.org/10.1007/s40092-014-0058-6

[20]. Takács, L. (1962). *Introduction to the theory of queues*. Oxford University Press.

[21]. Tao, L., Zhang, L., & Gao, S. (2014). M/M/R machine repair problem with spares and multiple modes of failure. *Mathematical Problems in Engineering*, *2014*, 1–10. https://doi.org/10.1155/2014/726071

[22]. Wang, K. H., Chen, W. L., & Yang, D. Y. (2009). Optimal management of the machine repair problem with working vacation: Newton's method. *Journal of Computational and Applied Mathematics*, *233*(2), 449–458. https://doi.org/10.1016/j.cam.2009.07.043

[23]. Yechiali, U. (2007). Queues with system disasters and impatient customers when system is down. *Queueing Systems*, *56*(3–4), 195–202. https://doi.org/10.1007/s11134-007-9031-z

[24]. Yang, D. Y., Wang, K. H., & Wu, C. H. (2010). Optimization and sensitivity analysis of controlling arrivals in the queueing system with single working vacation. *Journal of Computational and Applied Mathematics*, *234*(2), 545–556. https://doi.org/10.1016/j.cam.2009.12.046

[25]. Zhang, M., Hou, Z., & Wang, J. (2013). Performance analysis of M/G/1 queue with exponential working vacation and N-policy. *Applied Mathematical Modelling*, *37*(10–11), 6687–6699. https://doi.org/10.1016/j.apm.2013.01.048