High fidelity shelf stock monitoring
a framework for retail replenishment optimization

Author(s):
Metzger, Christian Peter

Publication Date:
2008

Permanent Link:
https://doi.org/10.3929/ethz-a-005627387

Rights / License:
In Copyright - Non-Commercial Use Permitted
High Fidelity Shelf Stock Monitoring –
A Framework for Retail Replenishment Optimization

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
Christian Peter Metzger
MSc in Electrical Engineering and Information Technology, ETH Zurich
born October 28, 1978
citizen of Maur, Zurich, Switzerland

accepted on the recommendation of
Prof. Dr. Elgar Fleisch, ETH Zurich and University of St. Gallen, examiner
Prof. Dr. Wolfgang Stölzle, University of St. Gallen, co-examiner
Dr. Stanley Gershwin, MIT, co-examiner

2008
Acknowledgments

First, I would like to thank Svetlena Taneva for her continuous support and encouragement. I am very grateful to have her in my life. I am also very thankful to my family for their love and support throughout the years.

For his great effort in guidance, constructive criticism, and excellent teaching, I would like to thank my thesis advisor, Professor Dr. Elgar Fleisch. His comprehensive knowledge of operations management and his enthusiasm for new and challenging projects are very inspiring to all his students. I am grateful that I had the opportunity to work with him.

For constant support and very helpful input I would like to offer special thanks to my co-examiners Professor Dr. Wolfgang Stölzle, University of St. Gallen, and Dr. Stanley Gershwin, Massachusetts Institute of Technology (MIT).

For their efforts in realizing my experimental setups, I would like to thank the people of the Soft Matter Physics Group at the University of Linz. Specifically, I would like to thank Professor Dr. Siegfried Bauer and Dr. Reinhard Schwödiauer for their enthusiasm and great support. In addition, I am grateful to Jan Meyer of the Wearable Computing Group at ETH for his help in building prototypes. Furthermore, I would like to express my sincere thanks to Frank Voermans of Sekisui Alveo who supported my ideas on weight-sensitive foams and who generously provided me with numerous foams for testing.

For his help in editing this document, I would like to thank David Nicholson.
## Table of Contents

### I Introduction
- I.1 Problem Statement ......................................................... 2
- I.2 Research Question and Methodologies ................................. 4
- I.3 Relevance ........................................................................... 8
- I.4 Thesis Outline ................................................................. 10

### II Review of the Retail Industry ............................................. 13
- II.1 The Retail Supply Chain .................................................... 14
- II.2 Retail Store Management ................................................. 20
- II.3 Retail Performance Measures and Drivers ......................... 25
- II.4 Out-of-stock ................................................................. 36

### III Continuous Shelf Inventory Monitoring .......................... 45
- III.1 Approaches to Shelf Inventory Monitoring ....................... 46
- III.2 RFID Technology .......................................................... 52
- III.3 Weight-sensitive Foam .................................................... 71
- III.4 A Qualitative Comparison ................................................ 91

### IV Mathematical Analysis .................................................. 103
- IV.1 Inventory Control Systems ............................................. 104
- IV.2 Shelf Inventory Models ................................................... 109
- IV.3 Shelf Inventory Models with Imperfect State Information .... 115
- IV.4 Evaluation ................................................................. 133

### V Conclusion ....................................................................... 165
- V.1 Key Findings ................................................................. 166
- V.2 Discussion of Research Findings ....................................... 172
- V.3 Theoretical Implications .................................................. 177
V.4 Practical Implications .................................................................178
V.5 Future Perspectives .................................................................179

APPENDIX A ......................................................................................184

APPENDIX B ......................................................................................190
List of Figures

Figure 1: Rapid prototyping as a methodology for instructional design.....................................6

Figure 2: Operations Research as a series of steps .................................................................7

Figure 3: Structure of the thesis .........................................................................................11

Figure 4: The generic flows between supply chain members ...........................................14

Figure 5: In general, there are three different ways for delivery, direct delivery, cross- docking, and delivery from a distribution center ............................................................15

Figure 6: The retail landscape ..............................................................................................18

Figure 7: A possible retail store layout with shelves in a repetitive pattern separated by aisles. ..................................................................................................................................21

Figure 8: Display of SKUs on a retail shelf .........................................................................21

Figure 9: Illustrates the different process steps for shelf replenishment and the activities associated with each step .......................................................................................24

Figure 10: Three levels of information sharing between organizations – transactional information, operational information, and strategic information ........................................29

Figure 11: Budget allocation for information technology in 2006 .........................................33

Figure 12: Diminution of product availability along the retail supply chain .......................37

Figure 13: The root causes of shelf unavailability ..................................................................43

Figure 14: The retail store is separated into two storage areas – backroom and sales floor ......48
Figure 15: The retail store’s inventory system distinguishes between backroom and individual product shelf space.

Figure 16: A component block diagram of an elementary feedback control loop.

Figure 17: Mass reading setup (portal, rotary or manual) to evaluate read rate performances.

Figure 18: Read rate performance on homogeneous pallets stacked with products containing liquids – read rate given for HF and UHF.

Figure 19: Read rate performance on homogeneous pallets stacked with metallic products – read rate given for HF and UHF.

Figure 20: The value of a transmitted byte is encoded in the 256 possible consecutive positions for a pulse.

Figure 21: Cross-section through the reader and tag antenna that are aligned on the coil axis. The tag antenna is tilted by the angle $\phi$.

Figure 22: Schematic of the tag with the analog front-end, logic unit, tilt sensor, and CMOS switch for data transmission.

Figure 23: The measurements show a nearly linear relation between the extracted peaks $u_0$ and the distance from the reader antenna.

Figure 24: Manchester coding with sub-carrier that represent a 1-bit and a 0-bit.

Figure 25: The plot of the scope values for $u_0$ at a distance of 43cm reveals the spikes (circled) occurring due to timing imprecision during Manchester encoding.

Figure 26: Novel’s AT-25A and Tekscan’s I-Scan and FlexiForce.
Figure 27:
The two electrodes enclosing a dielectric compressible material form a variable capacitor.................................................................76

Figure 28:
Experimental setup to measure the elastic modulus Y of the foam.........................78

Figure 29:
Illustration of strain vs. stress of polyolefin packaging foam in cross-section. The graph reveals a linear relationship between strain and stress at low stress levels of up to 0.5 bar; nonlinearities and hysteresis occur only at higher stress levels..........................................................................................................................79

Figure 30:
Illustration of small signal elastic modulus of the foam vs. pressure when pressure is increased up to 0.5 MPa at 0.56 MPa/min. Three distinct regions characterize the elastic behavior of the foam: linear (I) and nonlinear elasticity regions with softening (II) and densification (III) of the foam. .................................................80

Figure 31:
The cross-section electron microscope scan of polyolefin packaging foam shows few large and anisotropically shaped voids across the thickness of the foam...........................................................................................................................80

Figure 32:
The change in capacitance of a single sensor element correlates with the load........81

Figure 33:
The leads run in parallel across the foam and form capacitors at their crossover points. .................................................................................................................................81

Figure 34:
Shows the deviation in capacitance for TEE300.25 and TEE400.2 with a load of 11g on 10mm • 10mm electrode pads (left); and the deviation in capacitance according to a linear change in weight from 0-100g/cm² for TEE0300.2 on a 5mm • 5mm electrode pad (right). .............................................................................81

Figure 35:
Conductive silver leads run in parallel across the flexible polyolefin foam to form the 96 sensors. When multiplexed, the leads on different sides of the foam form the capacitive elements.................................................................83
Figure 36:
On the left: Products with highest out-of-stock rates are displayed on weight-sensitive foam. On the right: The graphical representation of loaded and unloaded capacitive elements that allow concluding the relative quantity of products on a shelf. ............................................................85

Figure 37:
Shows the existing inventory command (above) and the proposed command extensions (below). .................................................................98

Figure 38:
The display of the watchdog tag shows the decoded reader policy ID. ........................................99

Figure 39:
Shows the inventory level for periodic review and replenishment with different lead times. .................................................................109

Figure 40:
Shows the inventory level over time for one replenishment cycle with two different lead times. .................................................................114

Figure 41:
Illustrates the progression of actual inventory and recorded inventory over time and the measurement error that constitutes the difference between them. The upper line represents the actual inventory level and the lower one the recorded inventory level. At k=2, 4, and 5, a previously invisible item becomes visible while an item is removed. This reduces the measurement error by one for each occurrence. .................................................................120

Figure 42:
Illustrates the progression of actual inventory and recorded inventory over time and the measurement error that constitutes the difference between them as a superposition of false positives and false negatives. .............................................128

Figure 43:
Illustrates the progression of actual inventory (dashed) and recorded inventory (solid) over time. Once the recorded inventory reaches r = 1, replenishment is triggered and after a lead time, the actual inventory is raised to 11 and the recorded inventory to 9 to start a new cycle. .................................................................138

Figure 44:
Shows minimal costs for periodic review and RFID systems with respect to read rate and tag cost of $0.1 or $0.2 and review cost of $1, $0.5, $0.25 (λ = 10). ........................................................................................................143
Figure 45:
Shows the minimal operating costs for periodic review systems and RFID systems at different demand rates. For demand rates equal to and higher than $\lambda = 20$, RFID systems with $c_{tag} = $0.2 perform worse than periodic review, while RFID system with $c_{tag} = $0.1 perform better...........................................................146

Figure 46:
Shows the increase in minimal operating cost as the penalty cost increases (given for different read rates).................................................................................148

Figure 47:
Shows the reduction in units short with increasing penalty cost (given for different read rates)..................................................................................................148

Figure 48:
Illustrates the progression of actual inventory (dashed) and recorded inventory (solid) over time. The initial actual inventory level is 10. The initial recorded inventory level is 8 with initial measurement errors of 3 for false negatives and 1 for false positives..................................................................................................150

Figure 49:
Shows minimal costs as functions of detection rates for false negatives and false positives ($d_p$). ..................................................................................................152

Figure 50:
Shows the increase in minimal operating cost as the penalty cost increases (illustrated for different detection rates of false negatives and positives)...........157

Figure 51:
Shows the reduction in units short with increasing penalty cost (illustrated for different detection rates of false negatives and positives).................................157

Figure 52:
Shows a comparison of minimal operating cost for inventory management systems based on periodic review, RFID, and weight-sensitive foam (illustrated for different read rates and detection rates for false negatives and positives)........................................................................159

Figure 53:
Shows a comparison of minimal operating cost for inventory management systems based on periodic review, RFID, and weight-sensitive foam for $\lambda = 10$, 20, and 30 (illustrated for different read rates and detection rates for false negatives and positives).................................................................................160
Figure 54:
Shows the minimal cost for foam-based systems for different infrastructure cost in reference to RFID systems with tag costs of $0.1 and $0.2, respectively (demand rate $\lambda = 20$)..........................................................................................................................162

Figure 55:
Shows the minimal cost for foam-based systems for different infrastructure cost in reference to RFID systems with tag costs of $0.1 and $0.2, respectively (demand rate $\lambda = 20$)..........................................................................................................................162

Figure 56:
Shows the minimal cost for foam-based systems for different infrastructure cost in reference to RFID systems with tag costs of $0.1 and $0.2, respectively (demand rate $\lambda = 30$)..........................................................................................................................163

Figure 57:
Alveo’s Tee 0300.46. ..................................................................................................185

Figure 58:
Alveo’s Tee 0400.20. ..................................................................................................186

Figure 59:
Alveo’s TEE0400.20LT. ............................................................................................187

Figure 60:
Alveo’s TEE0400.48W. ............................................................................................188

Figure 61:
Alveo’s S604.25a. .....................................................................................................189

Figure 62:
Shows the minimal costs for RFID inventory systems with tag cost of $0.2 and different demand rates ($\lambda = 10, 20, 30$). ........................................................................193

Figure 63:
Shows the plots for $c_s$ and $p_c$ versus minimal cost and service level versus $p_c$......194

Figure 64:
Shows the different plots for $K_I$, $c_r$, $c_s$, and $\lambda$ versus minimal cost. For reasons of readability, only read rates of 100% and 90% are shown. The plots for lower read rates, however, show similar shapes.................................................................194

Figure 65:
Shows the different plots for $K_I$, $c_r$, $c_s$, and $\lambda$ versus minimal cost. For reasons of readability, only $d_n = 0.1$ and $d_p = 0$ are shown. The plots for lower read rates, however, show similar shapes.................................................................195
List of Tables

Table 1: Consumer reactions according to opportunity, substitution, and transaction costs ..........................................................40

Table 2: Energy consumption of the electronic components (measurements). .......................86

Table 3: Power consumption for different cycle times ..............................................................86

Table 4: Shows the weight of the products, the covered area on the foam, the footprint as seen by the sensors (actual pressure area), and the weight per area .........................87

Table 5: The table shows the false negatives and false positives for the tested products. ......88

Table 6: Shows in summary the advantages and disadvantages of RFID and weight-sensitive foam ..................................................................................................................102

Table 7: Shows minimal operating cost, units short per cycle, and corresponding service level for the optimal (r,S,T_r) policy for different λ (c_{rep} = 3, c_r = 0.5) .......................135

Table 8: Shows minimal operating cost, units short per cycle, and corresponding service level for the optimal (r,S,T_r) policy for different λ when maximum allocated shelf space is limited to 20 (c_{rep} = 3, c_r = 0.5). .........................................................135

Table 9: Shows minimal operating cost, units short per cycle, and corresponding service level for the optimal (r,S,T_r) policy for different c_{rep} and c_r (λ = 10, t_i = 0.02 days, T_h = 365 days, T_r = 1 day, c_s = $110, c_p = $10) .................................................................137

Table 10: Shows the minimal operating cost, units short per cycle, corresponding service level, and additional replenishments due to early triggering of the system for the optimal (r,S) policy at read rate φ, c_{tag} = $0.2, and λ = 10 (t_i = 0.02 days, T_h = 365 days, K_I = $12, c_{rep} = $3, c_s = $110, c_p = $10). .................................................................139
Table 11:
Shows the minimal operating cost, units short per cycle, corresponding service level, and additional replenishments due to early triggering of the system for the optimal (r,S) policy at read rate $\phi$, $c_{tag} = $0.1, and $\lambda = 10$ ( $t_l = 0.02$ days, $T_h = 365$ days, $K_I = $12, $c_{rep} = $3, $c_s = $110, $c_p = $10). ..............................................139

Table 12:
Shows the calculated values and simulation results for expected time until replenishment and expected number of units short for different (r,S) policies and read rates. Ten simulations are carried out and each runs for 10,000 consecutive days. ......................................................................................................142

Table 13:
Shows the minimal operating cost, units short per cycle, corresponding service level, and additional replenishments for the optimal (r,S) policy at read rate $\phi$ and $\lambda = 20$ ($c_{tag} = $0.2). ........................................................................................................144

Table 14:
Shows the minimal operating cost, units short per cycle, corresponding service level, and additional replenishments for the optimal (r,S) policy at read rate $\phi$ and $\lambda = 30$ ($c_{tag} = $0.2). ........................................................................................................144

Table 15:
Shows the minimal operating cost, units short per cycle, corresponding service level, and additional replenishments for the optimal (r,S) policy at read rate $\phi$ and $\lambda = 20$ ($c_{tag} = $0.1). ........................................................................................................145

Table 16:
Shows the minimal operating cost, units short per cycle, corresponding service level, and additional replenishments for the optimal (r,S) policy at read rate $\phi$ and $\lambda = 30$ ($c_{tag} = $0.1). ........................................................................................................145

Table 17:
Shows optimal policies for different detection rates of false negatives and positives, minimal operating cost, units short per cycle, and corresponding service level ($\lambda = 10$, $t_l = 0.02$ days, $T_h = 365$ days, $K_I = $7.5, $c_{rep} = $3, $c_s = $110, $c_p = $10). ........................................................................................................151

Table 18:
Shows the calculated values and simulation results for expected time until replenishment and expected number of units short for different (r,S) policies and detection rates of false negatives and positives. Ten simulations are carried out with simulation periods of 10,000 consecutive days..............................................154
Table 19:
Shows the optimal policy for different detection rates at demand rate $\lambda = 20$ as well as minimal cost, units short per cycle, and corresponding service level. ........155

Table 20:
Shows the optimal policy for different detection rates at demand rate $\lambda = 30$ as well as minimal cost, units short per cycle, and corresponding service level. ........155

Table 21:
Shows expected improvements in operating costs for the replacement of one system by another one (estimations for $\phi=1.0$, $d_n = 0.1$, and $d_p =0.1$). ...................160

Table 22:
Gives the infrastructure costs for foam-based systems at which these systems operate at lower cost than a comparable RFID system............................................161

Table 23:
Shows the simulation results for mean time until replenishment and units short for two sets of control parameters. Four simulations are carried out with a simulation period of 25,000 consecutive days each. ...............................................193
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Acceleration magnitude</td>
</tr>
<tr>
<td>APS</td>
<td>Advanced planning and scheduling</td>
</tr>
<tr>
<td>AC</td>
<td>Alternating current</td>
</tr>
<tr>
<td>ADC</td>
<td>Analog-to-digital converter</td>
</tr>
<tr>
<td>(\omega)</td>
<td>Angular frequency</td>
</tr>
<tr>
<td>T</td>
<td>Applied compressive stress</td>
</tr>
<tr>
<td>A</td>
<td>Area</td>
</tr>
<tr>
<td>Auto-ID</td>
<td>Automatic identification</td>
</tr>
<tr>
<td>C</td>
<td>Capacitance</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code division multiple access</td>
</tr>
<tr>
<td>CPFR</td>
<td>Collaborative planning, forecasting, and replenishment</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary metal-oxide semiconductor</td>
</tr>
<tr>
<td>CRP</td>
<td>Continuous replenishment</td>
</tr>
<tr>
<td>C</td>
<td>Coulomb</td>
</tr>
<tr>
<td>(\xi_k)</td>
<td>Cumulative of all items at state k that have become visible</td>
</tr>
<tr>
<td>(\xi_{nk})</td>
<td>Cumulative of the reduction in false negatives at state k</td>
</tr>
<tr>
<td>(\xi_{pk})</td>
<td>Cumulative of the reduction in false positives at state k</td>
</tr>
<tr>
<td>CRM</td>
<td>Customer relationship management</td>
</tr>
<tr>
<td>(T_c)</td>
<td>Cycle time</td>
</tr>
<tr>
<td>dB</td>
<td>Decibel</td>
</tr>
<tr>
<td>(\lambda)</td>
<td>Demand rate</td>
</tr>
<tr>
<td>(\varphi)</td>
<td>Detection rate</td>
</tr>
<tr>
<td>(d_n)</td>
<td>Detection rate of false negatives</td>
</tr>
<tr>
<td>(d_p)</td>
<td>Detection rate of false positives</td>
</tr>
<tr>
<td>DC</td>
<td>Direct current</td>
</tr>
<tr>
<td>d</td>
<td>Distance</td>
</tr>
<tr>
<td>EOQ</td>
<td>Economic Order Quantity</td>
</tr>
<tr>
<td>ECR</td>
<td>Efficient consumer response</td>
</tr>
<tr>
<td>(\zeta)</td>
<td>Electrical damping constant</td>
</tr>
<tr>
<td>EEPROM</td>
<td>Electrically erasable programmable read-only memory</td>
</tr>
<tr>
<td>EDI</td>
<td>Electronic data interchange</td>
</tr>
<tr>
<td>EPC</td>
<td>Electronic product code</td>
</tr>
<tr>
<td>EPCIS</td>
<td>Electronic product code identification service</td>
</tr>
<tr>
<td>E</td>
<td>Energy</td>
</tr>
<tr>
<td>ERP</td>
<td>Enterprise resource planning</td>
</tr>
<tr>
<td>EAN</td>
<td>European article number</td>
</tr>
<tr>
<td>ETSI</td>
<td>European telecommunications standards institute</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>$\varepsilon_i$</td>
<td>Event that at state $i$ an item becomes visible</td>
</tr>
<tr>
<td>$E[T_c]$</td>
<td>Expected cycle time</td>
</tr>
<tr>
<td>$E[n]$</td>
<td>Expected number of periods in one cycle</td>
</tr>
<tr>
<td>$E[y^-]$</td>
<td>Expected number of units short</td>
</tr>
<tr>
<td>$E[t_r]$</td>
<td>Expected time until replenishment</td>
</tr>
<tr>
<td>$fF$</td>
<td>Femtofarad</td>
</tr>
<tr>
<td>$f$</td>
<td>Frequency</td>
</tr>
<tr>
<td>FDMA</td>
<td>Frequency division multiple access</td>
</tr>
<tr>
<td>$g$</td>
<td>g-force ($9.80665\text{m/s}^2$)</td>
</tr>
<tr>
<td>GHz</td>
<td>Gigahertz</td>
</tr>
<tr>
<td>GPa</td>
<td>Gigapascal</td>
</tr>
<tr>
<td>GTIN</td>
<td>Global trade identification number</td>
</tr>
<tr>
<td>$g$</td>
<td>Gram</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross domestic product</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
</tr>
<tr>
<td>HF</td>
<td>High frequency</td>
</tr>
<tr>
<td>L</td>
<td>Inductance</td>
</tr>
<tr>
<td>IT</td>
<td>Information technology</td>
</tr>
<tr>
<td>$K_t$</td>
<td>Infrastructure cost</td>
</tr>
<tr>
<td>$\Delta_0$</td>
<td>Initial difference between actual and recorded inventory</td>
</tr>
<tr>
<td>$\Delta_n0$</td>
<td>Initial measurement error of false negatives</td>
</tr>
<tr>
<td>$\Delta_p0$</td>
<td>Initial measurement error of false positives</td>
</tr>
<tr>
<td>ICM</td>
<td>In-store customer marketing</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated circuit</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>J</td>
<td>Joule</td>
</tr>
<tr>
<td>KHz</td>
<td>Kilohertz</td>
</tr>
<tr>
<td>$t_l$</td>
<td>Lead time</td>
</tr>
<tr>
<td>LCD</td>
<td>Liquid crystal display</td>
</tr>
<tr>
<td>LAN</td>
<td>Local area network</td>
</tr>
<tr>
<td>$C(\ldots)$</td>
<td>Long-run average cost function</td>
</tr>
<tr>
<td>LF</td>
<td>Low frequency</td>
</tr>
<tr>
<td>m</td>
<td>Mass</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Mean</td>
</tr>
<tr>
<td>MHz</td>
<td>Megahertz</td>
</tr>
<tr>
<td>MPa</td>
<td>Megapascal</td>
</tr>
<tr>
<td>m</td>
<td>Meter</td>
</tr>
<tr>
<td>MEMS</td>
<td>Micro-electrical-mechanical systems</td>
</tr>
</tbody>
</table>
μm  Micrometer
μs  Microsecond
μW  Microwatts
mAh  Milliampere-hour
ms  Millisecond
nF  Nanofarad
k  Number of demands that have occurred
Nr  Number of replenishments
S  Order-up-to level
OOS  Out-of-stock
p, cp, pc  Penalty cost
ε  Permittivity
PDA  Personal digital assistant
π  Pi (3.14159)
T, Th  Planning horizon
POS  Point of sales
PE  Polyethylene
PP  Polypropylene
PCB  Printed circuit board
QR  Quick response
RF  Radio frequency
RFID  Radio frequency identification
y_{rk}  Recorded inventory level
c_{rep}  Replenishment cost
s, r  Replenishment threshold
R  Resistance
c_r  Review cost
T_r  Review period
SFA  Sales force automation
S_L  Service level
c_s  Shelf allocation cost
C_{Te(...)}  Single-period cost function
SDMA  Space division multiple access
σ  Standard deviation
SRAM  Static random access memory
SKU  Stock keeping unit
S  Strain
SCM  Supply chain management
$c_{\text{tag}}$ Tag cost

$\varepsilon_{ni}$ The event that at state $i$ the measurement error for false negatives has been reduced by one

$\varepsilon_{pi}$ The event that at state $i$ the measurement error for false positives has been reduced by one

$s$ Thickness

TDMA Time division multiple access

$\tau$ Time until replenishment

UHF Ultra high frequency

$\text{y}^{\prime}$ Units short

UPC Universal product code

VAN Value-added networks

VMI Vendor managed inventory

$V$ Voltage

$\lambda$ Wave length

WLAN Wireless local area network
Abstract

The highly competitive retail industry requires businesses to differentiate themselves through sensitive control of the managing elements. Despite a continuous effort to improve customer service levels, out-of-stock rates of retailers’ product assortment remain between 5-10%. A consumer facing an out-of-stock situation may buy a substitute, usually of a lower price to limit the risk associated with a mispurchase. Alternatively, the consumer may decide to buy that particular item or all items on the shopping list at another store. The combination of direct sales losses due to unavailability and indirect sales losses due to dwindling loyalty and unsatisfied consumers account for a sales diminution of up to 4%.

The root causes for out-of-stock situations at a retail store are manifold. The major causes result from insufficient in-store logistics. Infrequent manual shelf monitoring results in unnecessary and long-lasting out-of-stock situations. The disregard of shelf availability significantly affects the retail store’s performance. However, out-of-stock situations due to insufficient monitoring of shelf availability could be avoided because many of these products are located elsewhere in the store (e.g. in the backroom).

Retailers seek to improve the visibility of sales floor inventory by automating the monitoring process. High update rates on product quantities in combination with an immediate notification upon shelf depletion would allow for timely replenishments and a reduction in stock-out situations. The highly inaccurate nature of point of sales data generated at the check-out counter is insufficient for reducing out-of-stock situations. Therefore, retailers are currently evaluating other technologies that allow for automatic shelf monitoring. However, insufficient knowledge of a technology’s effect on the improvement of shelf inventory data and the operational costs associated with that technology makes the evaluation process challenging.

This thesis analyzes the retail shelf replenishment process and evaluates the effect of introducing shelf monitoring technology. Specifically, radio frequency identification (RFID) at item level and a novel design for the gravimetric detection of products on display are presented. RFID provides explicit identification of each individual item and significant value for the optimization of the entire supply chain. In addition, this thesis examines the benefits derived from the combination of RFID with sensors. These benefits include an increase in local resolution for improved shelf clustering and a technique to address privacy concerns associated with the use of RFID. In contrast to RFID, the application of weight-sensitive foam for gravimetric detection is limited to shelf inventory management. However, this limitation of weight-sensitive foam is compensated for by low manufacturing cost. In addition, weight-sensitive foam represents a stand-alone system that features a
simple installation since it does not require industry standards for proper operation. The advantages and disadvantages of both RFID at item level and weight-sensitive foam are outlined and developments for both technologies are described.

The replenishment process is modeled mathematically to evaluate the economic effects of these technologies on shelf inventory management. These models incorporate the shelf stock information generated by the monitoring technologies. Following a threshold-based replenishment policy, the inventory management system is optimized for each individual technology with respect to inaccuracies associated with measurement errors (e.g. read rates, false detection, etc.). The operating costs for an optimal shelf inventory management strategy are compared to a basic replenishment process that relies on manual inspection of product availability. A RFID-based inventory management system with tag cost of $0.1 shows a reduction of 10%-12% in operating costs in comparison to a periodic review policy. In general, RFID-based systems with tag cost of $0.2 do not yield an improvement. Shelf inventory management based on weight-sensitive foam may lead to a reduction in operating costs of 23% to 30% in comparison to a periodic review policy and of 13% to 22% in comparison to RFID. Incorporating technology specific costs and imperfect state information into the assessment of the economic impact on shelf inventory management provides a more sophisticated understanding of actual benefits.
Zusammenfassung

Aufgrund der starken Konkurrenz im Detailhandel müssen Detailisten versuchen, sich durch die behutsame Wahl und Kontrolle der Managementstrategie von ihren Wettbewerbern zu differenzieren. Trotz der kontinuierlichen Verbesserung des Kundenservice sind bei Detailisten nach wie vor etwa 5-10% des Produktsortimentes vergriffen. Konsumenten reagieren auf vergriffene Ware oftmals mit dem Kauf eines kostengünstigeren Substitutionsproduktes. Dadurch versuchen sie, das Risiko eines Fehlkaufes zu minimieren. Teilweise entschliessen sich Kunden auch dazu, einzelne oder alle Produkte des geplanten Einkaufes in einem Konkurrenzgeschäft zu erwerben. Die Kombination aus direktem Absatzverlust durch vergriffene Ware, wie auch indirekte Umsatzeinbussen durch schwindende Kundenloyalität, führen zu einem Rückgang des Gesamtumsatzes von bis zu 4%.


Die vorliegende Dissertation untersucht den Regalbefüllungsprozess im Detailhandel und evaluiert die Auswirkung einer automatischen Bestandsüberwachung auf die Kosten des Bestandsmanagements. Im Speziellen wird RFID (Identifikation durch Radiowellen) auf Produktebene sowie eine Neuentwicklung, die Produkte anhand ihres Gewichts identifiziert, vorgestellt und untersucht. RFID erlaubt einzelne Produkte eindeutig zu identifizieren, was nicht nur Vorteile für das Regalbestandsmanagement bietet, sondern auch

Um die ökonomischen Auswirkungen einer Technologie zur Regalbestandsüberwachung auf die Prozesskosten zu bestimmen, wurden in dieser Dissertation mathematische Modelle entwickelt. Diese Modelle beschreiben den Regalüberwachungsprozess wie auch den Prozess der Regalbefüllung und erlauben eine Optimierung unter Berücksichtigung der Informationsunsicherheit, welche durch ungenaue Messdaten entstehen kann (z.B. durch niedrige Leseraten oder Fehldetektionen). Die Prozesskosten, die mit der Bestandüberwachung durch eine jeweilige Technologie einhergehen, werden mit denen für die derzeit weitverbreitete manuelle und periodische Überprüfung der Regalbestände verglichen. Eine RFID-basierte Regalbestandsüberwachung mit Labelkosten von $0.1 führt zu einer Reduktion der Prozesskosten von 10%-12% gegenüber der manuellen, periodischen Überwachung. Sollten die Labelkosten jedoch $0.2 betragen, führt der automatisierte Prozess im Allgemeinen nicht zu einer Senkung der Prozesskosten. Durch ein Regalüberwachungssystem, das auf gewichtssensitivem Schaum basiert, können die Prozesskosten sogar um 23% bis 30% gesenkt werden, was gegenüber RFID einem Kostenvorteil von 13% bis 22% entspricht. Die Berücksichtigung von technologiespezifischen Kosten wie auch der Informationsunsicherheit durch Messfehler ermöglicht eine bessere Abschätzung der ökonomischen Vorteile, welche durch den Einsatz von Technologie zur automatischen Regalbestandüberwachung entstehen können.
This thesis seeks to improve the management of shelf inventory at a retail store. The first chapter gives a brief and general introduction into the research conducted in this thesis. It identifies the research problem, describes the scope of the thesis, presents the research methodologies applied, and summaries the theoretical and practical contributions.
I.1 Problem Statement

The trade in goods and services is as old as civilization itself. The retail business has grown to become an estimated $7 trillion industry worldwide [1]. In the US, the retail industry represents 15% of the gross domestic product (GDP), in Great Britain 11.7%, and 10.4% in Germany [2]. At the same time, the retail industry is among the most important employers in these countries (idem).

The core business of the retail industry consists of selling goods in small quantities to the general public. Hence, a retailer seeks to meet customer demand with a concerted product assortment, competitive prices, and high product availability. Despite the highly competitive industry, significant developments in retail technology, and retailers’ continuous striving to improve customer service, the average rates on product availability are still around 90-95% (cf. II.4). This rate for product availability corresponds to a stock-out rate of 5-10% which results in retail losses of up to 4% (idem). In addition to direct sales losses due to product unavailability, retailers may also incur indirect sales losses due to dwindling loyalty of unsatisfied consumers. In metropolitan areas, competition is high and consumers are offered a large variety of brands and stores. High accessibility to retail stores keeps the store switching costs low. Therefore, retailers are obliged to provide high product availability to achieve satisfaction and loyalty among customers. The fact that product availability is a critical performance measure raises the question: why do retailers fail to significantly reduce out-of-stock rates?

The majority of out-of-stock situations at the retail store are a direct result of inefficient retail store practices. About one fourth to one third of stock-out situations is due to inefficient replenishment practices where products are at the store but not on the shelf, and consequently, these products are not available for sale (cf. II.4).

The numerous studies, conducted around the world, that identified retail out-of-stock situations as a significant problem were carried out through manual monitoring of specific product categories at different stores (cf. II.4). These studies did not rely on in-store inventory data due to its highly inaccurate nature (cf. II.3.3). The lack of inventory visibility represents a major cause for inefficiencies in in-store logistics. Retail store managers are often not aware of empty shelves and the quantity of remaining stocks. Therefore, sales floor staff cannot be directed to shelves that need to be replenished. Even if retailers’ awareness of out-of-stock situations was raised, how could they address this issue?
Currently, most retailers pursue a response-based logistics strategy and replenish their shelves from the backroom (cf. II.2.3). The sales floor staff manually inspects the shelves and estimates the remaining stock in the backroom based on point of sales (POS) data. The POS data is generated by scanning a product’s barcode at the check-out, a data acquisition process that is highly susceptible to errors (cf. II.3.3). Manual shelf stock inspection is also error prone and occurs at low cycle time due to significant labor cost for each inspection. Consequently, store managers have to make replenishment and ordering decisions based on inaccurate inventory data that fails to differentiate between backroom and sales floor.

Retailers seek to improve the visibility of sales floor inventory. The automation of inventory audits allows for the elimination of errors introduced by the manual execution of this process, reduces the costs of data collection and allows for continuous updates. However, could automatic inventory monitoring on retail shelves, which generates a replenishment alert if the inventory drops below a certain threshold value, eliminate out-of-shelf situations?, and which technology meets the requirements for this application best?

Radio frequency identification (RFID) has received increased public awareness since a consortium of companies has investigated its use for supply chain optimization in the late 1990’s. The increased interest in RFID resulted from Wal-Mart’s request in 2003 that the top 100 suppliers use RFID at the pallet and case level by 2005 [3]. The utilization of RFID suggests to significantly improving a supply chain’s visibility, which will support the management of supply chain processes and increase the supply chain’s efficiency. Gartner Research accredits RFID a hype cycle. After the “peak of inflated expectations” was passed in 2005, the interest in RFID has been declining, approaching the “trough of disillusionment”, after which it will increase again [4]. In contrast to Gartner’s forecast, Frost & Sullivan expect an annual growth rate of 32% for the RFID-based application market in the period from 2003 to 2010 [5]. Their stated reasons for this growth include the development and adoption of standards, falling prices of RFID equipment, growing support for RFID-based supply chain management applications, technology innovations, increasing end-user awareness, and the introduction of RFID into new applications (idem). What process improvements can be expected from the introduction of RFID into the shelf replenishment process considering an adoption of RFID at item level?

While RFID represents an object-centered technology, where objects are detected and identified through intermediate devices, alternative technologies for inventory monitoring exist that are permanently deployed in the environment of the retail store (cf. III.3). Which technology from this group of environment-embedded
technologies shows potential to support the replenishment process, and how does it compare to RFID? What is the economic impact of shelf stock monitoring on the inventory management costs?

Despite the many new ideas and technologies available, many retailers have failed to improve retail store performance. The reason is that retailers lack a framework for informed conclusions on the specific performance of a technology [6]. In what follows, the effect of automatic shelf stock monitoring on replenishment process costs and on out-of-stock rates is analyzed along with a comparison and evaluation of different technologies. The findings intend to support retailers in the decision making of introducing a specific technology to improve the replenishment process.

1.2 Research Question and Methodologies

1.2.1 Research question

This thesis addresses the issue of improving retail shelf inventory management through the introduction of technology by raising the following research question:

*How can continuous shelf stock monitoring at a retail store improve shelf inventory management with respect to minimal total process costs?*

In order to structure the approach of answering this research question, the following sub-questions will be addressed:

- What technologies show potential to support shelf stock monitoring?
- What are the properties of these technologies?
- How can a mathematical model of a shelf inventory system account for imperfect state information?
- How do optimized (r,S) replenishment policies for different technologies compare?

Answering the main research question will prove the following hypothesis: *Automatic monitoring of shelf stocks based on RFID or weight-sensitive foam has an effect (positive or negative) on the total retail shelf inventory management costs.*
Accordingly, the null hypothesis is: *Automatic monitoring of shelf stocks does not have any effect on the total retail shelf inventory management costs.*

This research assumes a rational approach towards the adoption of new technology, where a technology is only adopted if the expected cost benefits exceed the ones possibly derived from alternatives. The research question stated above is derived from an actual challenge that retailers face. By answering the research question within a business perspective, the thesis intends to generate normative conclusions and guidance for decision makers.

This thesis aims to increase the knowledge on technology that may support the retail shelf replenishment process and analyzes two different groups of technologies – object-centered and environment-embedded. A theoretical framework and the development of formal models help to quantify the value of these technologies based on a total process cost function. In addition, these models allow comparing their value for the improvement of the replenishment process. Consequently, this framework may provide practical value as it increases the understanding of different technologies and their individual benefits for retail shelf inventory management.

RFID is an evolved technology and a prominent representative of object-centered technologies. However, there is no protruding technology for the group of technologies that may be embedded in the environment of a retail store. Despite the fact that RFID is given credit for substantial improvements in supply chains, it also has several disadvantages such as high costs at item level, the requirement of product modification (equipping each item with a RFID tag), privacy issues, etc. (cf. III.2). Therefore, RFID is compared to a proprietary solution that is specifically designed to meet the requirements for shelf inventory management at a retail store. This solution detects products through their physical characteristics and does not require any product modifications.

**I.2.2 Research methodologies**

**I.2.2.1 Rapid prototyping**

The lack of an applicable environment-embedded technology leads to the development and evaluation of weight-sensitive foam (cf. III.3). This thesis follows Tripp and Bichelmeyer’s understanding of a methodology for rapid prototyping [7]. Rapid prototyping forms an instructional design strategy based on five stages that may not necessarily be conducted in a consecutive manner but could be conducted simultaneously (Figure 1). These stages are: (1) *assess needs and analyze content*, (2) *set objectives*, (3) *construct prototype*, (4) *utilize prototype*, and (5) *install and...*
maintain system. The first and second stages offer a plan of action that leads to the construction of a prototype with a strong focus on immediate problem solving. Through the utilization of the prototype, strengths and weaknesses are identified. These findings form the input for further developments, and the development process starts again at the stage of needs and objectives. The prototyping process is completed once the prototype meets the requirements of the application. Rapid prototyping is effective in situations where synthesis and modifications must occur quickly and the method cannot be applied in a linear manner.

![Figure 1: Rapid prototyping as a methodology for instructional design](image)

I.2.2.2 Mathematic modeling

Mathematic modeling along with statistics and algorithms is a scientific method used in operations research or operational research to support decision making between various courses of action available to accomplish specified objectives. Problems concerned with the improvement or optimization of operations coordination and execution are addressed with operations research. Some of the tools used by operations researchers are statistics, optimization, stochastic, queuing theory, game theory, linear programming, simulation, and Markov processes [8]. Operations research intends to provide a scientific approach to derive a best solution to a complex problem (process illustrated in Figure 2).
First, the problem is formulated, which requires the understanding of the organizational climate, expectations and alternative causes of action [8]. Next, the problem is mapped with a model that represents the system, processes and relationships with equations and formulas. In order to test and operate the model, sufficient and adequate input data must be available. Updates and modifications to the model provide a tentative solution to the problem. The tentative solution requires validation to ensure that the model provides reliable predictions of the system’s performance and be applicable over time. Finally, the solution is implemented.

This thesis is concerned with the analysis and improvement of retail shelf inventory management costs. Most mathematical inventory models are designed to address the two fundamental issues of when a replenishment order should be placed and how much should be ordered. The complexity of the model depends on the assumption that one makes about the demand, cost structure, and physical characteristics of the system. The objective of virtually all inventory control models is to minimize costs [10]. Consequently, an inventory control model will be optimized for lowest total process costs.
I.2.3 Limitations

The research procedure applied in this thesis follows the research methodologies described above. The research methodology on rapid prototyping strongly suggests that the findings derived from the evaluation of demonstrators and prototypes provide a thorough understanding of the effect of such hardware and software on the application under research. Nevertheless, the results of the evaluation are limited to the applied scope of the research and cannot capture events occurring outside that scope. Although the evaluation of a demonstrator or prototype can prove its feasibility for a certain application, only a large scale deployment provides hard evidence that the expected results can be realized.

Similar limitations apply to mathematical models intended to describe an actual process. The modeling of a process requires a certain level of abstraction that disregards factors unrelated to the outcome. This abstraction reduces complexity making the findings meaningful and useful. However, disregarding certain factors based on informed decisions still carries the risk of neglecting important input information.

I.3 Relevance

I.3.1 Theoretical relevance

As stated above, most retail stores inspect inventory levels on retail shelves periodically. The periodic reviewing of inventory levels can be represented by a stochastic model with periodic review (e.g. the AHM-Model [11]). This model serves as benchmark to measure process improvements through the use of continuous inventory monitoring. Although stochastic models with continuous review have been extensively researched (cf. IV), the effects of incorporating such monitoring technology on total process cost have received little attention. Additionally, most current stochastic models with continuous review do not address the impact of imperfect operation of such technology on inventory records. While it is assumed that automatic inventory monitoring provides accurate inventory records, the utilization of such technology itself introduces errors.

This thesis analyzes the effects of inventory monitoring systems on total shelf inventory management costs. Additionally, it examines the reliability of automatic monitoring technology, specifically RFID and weight-sensitive foam, and analyses the effects of inaccurate data on the replenishment process. For that purpose, the model for continuous review is extended in order to account for such technology. Based on this model, optimal replenishment strategies under the assumption of
imperfect operation of a specific technology are derived. In addition, the development of a model that accounts for the use of different technologies allows comparing the optimized replenishment process cost functions.

In order to obtain a model that forms a suitable representation on how automatic monitoring technology affects the replenishment process, the properties of such technology have to be studied extensively. Therefore, the potential benefits and challenges of RFID at item level are analyzed. In addition, the impact of an increased level of detail on location information derived from sensors is evaluated. Weight-sensitive foam is developed due to the lack of an existing technology that can be permanently deployed to retail shelves. The design and implementation of a proprietary solution with respect to minimal manufacturing and system maintenance costs allows comparing two different approaches for automatic shelf inventory monitoring. While RFID provides a specific level of detail, the information granularity derived from weight-sensitive foam is variable and reflected by manufacturing costs.

Consequently, the main contribution of this thesis is a quantitative and qualitative analysis of the impact of different technologies on retail shelf inventory management and the development of inventory models that account for errors induced by the monitoring technology. The impact of each technology on the total process costs is compared to other technologies and to the basic approach of manual inventory inspection in a periodic manner.

I.3.2 Practical relevance

The awareness that poor replenishment strategies lead to extensive stock-out situations and significant sales losses increases among retailers. However, retailers have insufficient knowledge about opportunities that will effectively improve shelf inventory data and decrease stock-out situations and sales losses. This thesis evaluates technologies to improve the replenishment process, details the advantages and challenges of deploying a specific technology, and discusses the possible benefits.

In addition, this thesis provides a stocking system that detects merchandise through a gravimetric sensor array to derive shelf stocks. This system, which is specifically designed for the shelf replenishment application and accounts for low manufacturing costs, offers an alternative to RFID. The proprietary system suggests providing sufficient information about shelf stocks and low adoption barriers.
1.4 Thesis Outline

Chapter two describes the retail supply chain – the flow of goods from the manufacturer to the retail store. The retail business and its value proposition are explained to backup the problem statement formulated above. Replenishment practices are examined along with material handling activities at different locations at the retail store. In addition, a detailed analysis of out-of-stock situations, their causes and extent, and the financial consequences for a retail store. The influence of information technology and information sharing on supply chain performance is studied as well as different levels of information sharing between organizations, investments made in information technology, and data accuracy at the retail store.

Chapter three describes different approaches to automatic shelf inventory monitoring and explains how the level of information granularity may depend on the level of technology adoption at the retail store. This chapter groups potential technologies into object-centered and environment-embedded technologies and compares the individual advantages and challenges. Specifically, this chapter analyzes the impact of RFID, an object-centered technology, on the retail sector and takes into account that the data accuracy significantly depends on the read rates. This is followed by the examination on how the combination of RFID and sensors could compensate for several shortcomings of RFID. The chapter continues with the introduction of weight-sensitive foam, an environment-embedded technology. The properties of such foam are derived and the design of a weight-receptive sensor array system is presented. A detailed evaluation of the system shows its feasibility for the shelf replenishment application. The chapter concludes with a qualitative comparison of the two technologies.

In chapter four, the quantitative benefits of introducing automatic shelf inventory monitoring technology are evaluated. Models for different technologies and levels of technology adoption are developed. The basic model represents the current replenishment strategy where store personnel periodically inspect shelf stocks. Incorporating inventory monitoring technology leads to continuous models that must account for imperfect state information due to limited operational accuracy. Hence, the effect of each technology on shelf inventory management costs is obtained.

Chapter five concludes the thesis with a summary of the major findings and discusses the theoretical and practical implications. The chapter ends by listing future perspectives. Figure 3 illustrates the structure of the thesis.
### Chapter 1: Introduction

Problem statement  
Research question and approach  
Contribution

### Chapter 2: Review of the Retail Industry

<table>
<thead>
<tr>
<th>Content</th>
<th>Describes the retail industry, operations at the retail store, out-of-stock as a performance indicator and the importance of IT.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result</td>
<td>A detailed understanding of activities that result in OOS situations.</td>
</tr>
</tbody>
</table>

### Chapter 3: Continuous Shelf Inventory Monitoring

<table>
<thead>
<tr>
<th>Content</th>
<th>Describes, evaluates, and develops different technologies that show potential to support the replenishment process.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result</td>
<td>A detailed understanding of the advantages and challenges of different technology.</td>
</tr>
</tbody>
</table>

### Chapter 4: Mathematical Analysis

<table>
<thead>
<tr>
<th>Content</th>
<th>Develops mathematical models for shelf inventory management that account for imperfect state information.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result</td>
<td>A detailed understanding of the performance of each technology based on a process cost function.</td>
</tr>
</tbody>
</table>

### Chapter 5: Conclusion

| Summary of major findings  
|----------------------------|
| Practical and theoretical implications  
| Future prospects |

**Figure 3: Structure of the thesis**
Review of the Retail Industry

The retail industry is characterized by the overall effort of bringing merchandise from the manufacturer to the consumer. This chapter describes the retail supply chain, classifies the shelf replenishment process, examines the material handling activities involved, and analyzes current shortcomings as well as their consequences for the retail business. The identification of shortcomings reveals potential for the introduction of technology to support shelf inventory management.
II.1 The Retail Supply Chain

II.1.1 Structure of the retail supply chain

Providing products to a customer involves one or more entities. The activities of the entities include manufacturing, assembling, distributing, transporting, and selling a product to a customer. A network of entities directly involved in the upstream or downstream flow of initial raw materials, services, finances, and/or information to a customer constitutes a retail supply chain [12]. Stern et al. describes nine generic flows between partners in a supply chain – physical possession, ownership, promotion, orders and payments, negotiation, financing, risking, and information [13] as illustrated in Figure 4. Supply chains contain one or more entities. Likewise entities may participate in one or more supply chains. Instead of owning the entire value chain, companies usually cooperate in supply chain networks in order to benefit from partners that specialize in domains outside the company’s core business.

![Diagram of the generic flows between supply chain partners](image)

Figure 4: The generic flows between supply chain partners [13].

A variety of sources exist for customers to satisfy their demands. Therefore, companies place significant importance on maximizing customer accessibility with minimum costs. Previously, ensuring high product availability meant maintaining inventory throughout the supply chain. However, holding substantial inventory carries the risk of a decline in value if the buying habits of customers change and the demand for a product begins to wane. In addition, holding extensive inventory results in significant fixed capital. Therefore, companies across the entire supply chain coordinate their activities by sharing information on supply and demand. Communication within the supply chain seeks to increase competitiveness and provide the lowest prices to consumers.
In general, retailers receive deliveries in three ways – direct delivery, cross-docking, and deliveries from the manufacturer or from a distribution center owned by the retailer. A distribution center serves to sort, consolidate, and store material for scheduled outbound shipping. Cross-docking represents an intermediate form where products are sorted, consolidated, processed, and labeled but not stored at a specific docking point. Inventory is avoided by synchronizing the receiving of goods with delivery to retail stores.

![Diagram showing delivery methods](image)

**Figure 5:** In general, there are three different ways for delivery, direct delivery, cross-docking, and delivery from a distribution center (adapted from [14])

In addition to the supply chain a retailer participates in a distribution channel or marketing channel [15]. Sales, negotiation, and ordering are done within the distribution channel. According to Hugos et al., the distribution channel allows breaking up the buying and selling process as well as the related negotiation into manageable tasks. Additionally, distribution channels influence the natural ebbs and flows of a supply chain [15]. Holding inventory along the supply chain can even fluctuating demands.

In summary, a general definition of a supply chain is given according to Lummus et al. as follows:

*All the activities involved in delivering a product from raw material through to the customer including sourcing raw material and parts, manufacturing and*
assembly, warehousing, and inventory tracking, order entry and order management, distribution across all channels, delivery to the customer, and the information systems necessary to monitor all of these activities [16].

II.1.2 Supply chain management

The management of a supply chain is the coordination and integration of all activities related to a supply chain into a seamless process. However, the definition of supply chain management (SCM) differs across authors. Mentzer et al. classify the definitions into three categories: a management philosophy, implementation of a management philosophy, and a set of management processes [12]. Cooper et al. describe SCM as an integrative philosophy to manage the flow of goods of a distribution channel [17], which was then extended into a concept of a multiform effort to manage the total flow of goods. Hence, supply chain management as a philosophy understands that each firm has a direct or indirect influence on the performance of all other participating firms [18]. As a management philosophy, supply chain management takes an integrative view at the supply chain emphasizing cooperative efforts. These efforts serve to optimize inter-firm partnerships and operational processes aimed to create customer value [12]. Other authors understand supply chain management as a set of activities to implement a management philosophy by focusing on the different activities and their embodiments within a supply chain [19], [20], [18].

Lambert et al. identifies supply chain management as a new way of managing the business and its relationships, and defines supply chain management as:

the integration of key business processes from end user through original suppliers that provide products, services, and information that add value for customers and other stakeholders [21].

With this definition Lambert et al. provide an understanding of the term supply chain management that focuses on management processes instead of activities. Davenport and Ross support this view of supply chain management by defining processes as the activities of those entities that determine the actual flow of goods and services to the market [22], [23]. To successfully implement SCM, all functions within a supply chain must be reorganized as key processes [24]. Mentzer et al. summarize the critical differences between the traditional functions and the process approach and isolate meeting the expectations of the customer as a crucial distinction [12].

The retail industry represents a highly competitive industry characterized by low margins and high sales volumes [25]. The core of the retail business operates by breaking up bulk products into smaller lot-size quantities in order to sell them at
shopping facility that are conveniently accessible to customers. Retailers seek to combine large product variety and high availability. Supply chain management aims to establish a profitable and sustainable position against the competition while providing high customer service. Customer service improves by lowering the costs, increasing stock availability, and reducing order cycle time [12]. In order to improve a firm’s competitive advantage and profitability through supply chain management, competition should be understood as rivalry among supply chains rather than among companies [26]. Therefore, supply chain management is concerned with improving efficiency in a strategic context to obtain competitive advantage that ultimately bring profitability [12].

II.1.3 Logistics management

Lambert et al. states that most practitioners, consultants, and academics had not drawn a sharp line between supply chain management and logistics management [21]. Logistics management was defined by the Council of Logistics in 1986 as:

[…] the process of planning, implementing, and controlling the efficient, cost-effective flow and storage of raw materials, in-process inventory, finished goods, and related information flows from point-of-origin to point-of-consumption for the purpose of conforming to customer requirement.

Lambert et al. explains the confusion by noting that logistics is understood as a functional area within a company and as broader concept that deals with the management of material and information flows across the supply chain. In the modified declaration from the Council of Logistics in 1998, logistics management is explicitly defined as only a part of SCM:

Logistics is that part of the supply chain process that plans, implements, and controls the efficient, effective flow and storage of goods, services, and related information from the point-of-origin to the point-of-consumption in order to meet the customers’ requirements [21].

This definition locates logistics as one of the functions contained within supply chain management. In addition, the Council of Logistics’ definition aligns with Giunipero et al. who emphasize efficient movement and storage within the supply chain process to fulfill customers’ requirements [27].

II.1.4 Retail segments

According to the Oxford English Dictionary, a retail dealer or trader is an entity that sells goods in small quantities to the general public. Conceptually, retail divides into two forms - functional and institutional. In functional form, retail stands for the balancing of shortage differences between manufacturing and demand [28]. In its
institutional form, retail comprises the reselling of goods to the end consumer without any processing or refinement [28]. The retailer represents the final tier in a distribution channel, selling products directly to the end user or consumer [15].

The retail category contains four segments; retailers seeking premium offerings by directing their strategies toward quality-conscious markets, retailers that appeal to price-conscious markets, retailers that thrive because of their value offering, and retailers unable to provide high levels of value relative to their competitors [29], [30]. Most retailers belong either to the low price or the value offering segment. Retailers in the innovative segment, unable to keep pace with competition, usually transition toward the value offering segment. Sometimes those in the innovative segment move towards the low price segment or exit the retail landscape (Figure 6).

![Figure 6: The retail landscape (adapted from [29]).](image)

According to Grewal et al., the low price segment experienced double-digit growth in store openings. This finding suggests that a good business model is one that contains fast-turnaround items. Grewal et al. note that although dollar store transactions are minimal with only $9, the average margin of 32% outpaces those of supermarkets with 31%, convenience stores with 29%, drug stores with 27%, discounters with 24%, and warehouse clubs with 11% [30]. Low operation cost and high product availability provide good customer service and competitive
profitability. Usually, competitive retailers in the value offering segment have successfully transformed from an innovative or low price retailer into a retailer that serves the needs of a large number of customers. In order to keep their customer base, these retailers are receptive to changes in the customer’s buying behavior. In this way, the retailers detect changes in product demand and react with adjusted offerings to continuously meet the customer requirements.

In order to become or remain competitive, a retailer focuses either on optimization or strategy [31]. If the retailer focuses on strategy, Porter recommends uniqueness in the value proposition (idem). Except for requiring uniqueness in the value proposition, Grewal et al. support a similar understanding of retailer competitiveness. Grewal et al. do cite the importance of differentiating one’s business from the competition by sensitive control of the managing elements. As examples, Grewal et al. offer store appearance, service level, merchandise, price, supply chain optimization, and technology.

A retailer should comfort its customers by making shopping an enjoyable experience. Ensuring an enjoyable customer experience requires an atmosphere that serves to extend the amount of time the customer spends in the store. In addition, the retailer needs to offer good customer service. In particular, store personnel need to be able to provide decision convenience (appropriate product information), access convenience (find merchandise within the store), transaction convenience (facilitate check-out and return), benefits convenience (help the customer to understand the benefits of a product), and post-benefit convenience (rectify post-purchase problems) [30].

Another key element in creating value for the customer is effective merchandise management. Combining large product variety and/or unique merchandise with high service levels enables a business to distinguish itself from the competition. The large product variety serves the customers’ preference to do all their shopping at one store [32].

The sales price represents the only managing element that creates monetary revenue for the retailer. The sales price is also a critical factor in the customer’s decision to buy or not to buy. The customer always compares the benefits accruing through the purchase with its sacrifices. The more one of them outweighs the other, the easier it is for the customer to make a purchase decision – positive or negative. In order to offer best purchase value for its customers a retailer must efficiently manage its supply chain. The use of technology to collect and share data supports the successful management of a supply chain. While most retailers collect sales data at the point of sale for planning, forecasting, ordering, inventory management, and
replenishment, only some retailers share their data with their supply chain partners for production and distribution planning. Sharing sales data with all supply chain members may increase a supply chain’s efficiency and generate additional profit.

II.2 Retail Store Management

The retail store represents the last tier in the retail supply chain. Effective in-store practices are paramount in order to provide high service levels to the customer. In addition, effective in-store practices influence the overall performance of the entire supply chain. Product availability deteriorates along the supply chain [33], [34]. Therefore, the level of product availability at the store is significantly lower than at other echelons along the supply chain. Low service levels at the retail store are mainly caused by inefficient retail processes [35]. This subsection describes the inside of a retail store, its replenishment practices and handling operations and details the cost structure of the in-store processes.

II.2.1 Store layout

The layout of a retail store usually describes a repetitive pattern of tiers of shelves, bins, or pegs separated by aisles so the customer can easily reach the merchandise (Figure 7). Each SKU in the assortment of a retailer gets one or more slots on a shelf, bin, or peg. The slot of a SKU is defined by its location and space allocation. The location of the SKU is determined by the assigned shelf and the SKU’s position on the shelf. Shelves are divided into facings that describe the one side of a product’s packaging that “faces” the customer. Therefore, each facing carries only one SKU (Figure 8). The number of facings per SKU determines the shelf space that needs to be allocated for a certain SKU. The number of items per SKU on the shelf is determined by the number of facings and the depth of the retail shelf. The different outlines of SKUs result in varying slot sizes and total allocated shelf space.
The interrelationship between inventory management and retail shelf management has garnered significant attention. It is a well-established phenomenon that high product availability positively influences the sales of many items [37], [38], and that the sales depend on the number of items displayed [39]. Due to shelf space limitations, efficient shelf space allocation management and product assortment are critical retail operations. Many manufacturers are willing to pay premiums in order to obtain preferred retail locations [40]. In return, manufacturers expect retailers to efficiently manage on-shelf inventory to provide high availability of the manufacturer’s products. Therefore, excellent control of shelf availability is essential for store efficiency and profitability.
II.2.2 Incoming goods

A retail store usually receives deliveries from the distribution center at the delivery dock of the backroom facility where the shipping data is inputted into the inventory system. Most products are delivered in stackable cases and boxes on pallets. This delivery method allows for space efficient storing in the backroom where goods are typically stacked up high and close together. The backroom serves as a buffer between the delivery period and the actual replenishment of the shelves thereby smoothing the fluctuation in demand and lead time. Order batching allows placing large orders that reduce ordering costs due to full truck load economies. The stored ordering lot size is usually broken into smaller sizes before being moved onto the sales floor. Backroom storage is specifically required for goods with high demand that are usually ordered in large quantities.

II.2.3 Replenishment strategies

According to Wong et al., there is a growing trend towards Direct-To-Store and One-Touch-Replenishment store replenishment practices that reduces the importance of the backroom [34]. However, the best position of inventory within the distribution system is next to the customer [41]. Due to storage space limitations on the sales floor, a considerable amount of inventory is stored in the backroom. Therefore, retail shelves are usually replenished directly from the backroom. Traditional replenishment strategies are based on either anticipatory logistics (push policy) or on response-based logistics (pull policy). Anticipatory logistics starts from the backroom where stockers manually inspect the stock quantities. In combination with sales data (POS data), the remaining stocks on the shelves is estimated. The cycle time for replenishments is determined by the sequence of arrangement in the backroom instead of low stock quantities on shelves [34].

Response-based logistics requires the staff to manually inspect the retail shelf for product availability and compare the remaining stock against the allocated shelf space. Product types that are below a certain quantity but estimated to be available in the backroom, are added to a picking list either through manual recording or through electronic scanning of the product identifier. This replenishment policy is responsive to low stock levels on shelves.

The factors influencing the effectiveness of current replenishment polices are observation delay, check-out delay, picking list generation, localization of products in the backroom or on the shelf, the physical movement of products from the backroom onto the sales floor, maximum shelf space allocation, and the product removal pattern [34]. The observation delay describes the time span between the
moment the quantity of a product falls below a threshold and its detection by the staff. Low observation frequency may cause delays in the detection of empty shelves. The check-out delay describes the time that elapses between the removal of an item from the shelf and its scanning at the check-out. Inventory estimations that are based on POS data may show inaccuracies due to check-out delays. The picking list records the quantity that needs to be replenished. However, if the replenishment requires a significant amount of time and the product removal rate is high, the picking list generation may be obsolete. For timely replenishment, it is critical to efficiently locate products and move them from the backroom to the shelves. The shelf space allocation not only affects the product availability and influences the purchasing behavior of customers but also affects the number of observation and replenishment cycles required for a certain product removal pattern (demand).

II.2.4 Material handling activities for shelf stacking

The replenishment of retail shelves is a very labor intensive, daily task. Empirical studies suggest that material handling costs in the retail supply chain represents a significant share of the total operational costs [42], [36]. The activity of moving new deliveries from the delivery dock or stored material from the backroom onto the sales floor constitutes a significant work load on the store clerk. The extensive labor time required to fulfill this task results in high operational costs.

The handling activities that compose the shelf replenishment process divide into several subtasks. Prior to shelf stacking, products are moved to the sales floor and to the shelf that requires replenishment. Next, the store clerk breaks the case packs into smaller consumer units and positions them on the shelf that needs to be replenished. In order to increase customer awareness and promote purchase, the clerk arranges the display of products according to store policy. The maintenance of the shelf requires the following: the rearrangement of products according to their expiration dates (products with an earlier expiration date are usually moved to the front where they are easier reachable), and the removal of expired products or old inventory if necessary. These steps are necessary before the shelf can be stacked with new items. The stacking of a shelf is completed with the disposal of the empty case packs (Figure 9).
Figure 9: Illustrates the different process steps for shelf replenishment and the activities associated with each step.

Cersue et al. conducted an empirical study on stacking time and broke down the shelf stacking activities into grabbing and opening a case pack, searching for the assigned location, walking to the assigned location, preparing the shelf for stacking the new items, filling new inventory on the shelves, filling the old inventory back on the shelves, and disposing the waste package [43] (see also [44]) (Figure 9). They state that the total average stacking time per order line is about 57s with a standard deviation of 36.6s. Inventory stacking requires the most amount of time (48%), followed by grabbing and unpacking the case (20%), disposing the waste (13%), walking to the shelf (8%), preparing the shelf (6%), searching for the location (4%), and by stacking the old inventory (1%). Kotzab et al. examined the daily amount of time spent on replenishments at smaller-sized retail stores (~750SKUs) in Austria. Although there were significant differences between store formats, the mean value for shelf replenishment is given as 6.4h/day [2]. However, this process ties up a significant amount of employee hours (14.3h/day) to it.

II.2.5 Cost structure

A typical American supermarket carries about 45,000 different items or stock keeping units (SKU) on an everyday basis. Store occupancy costs range from about $20 per square foot (about $214 per sq meter) for dry grocery shelf space, over $50 per sq ft (about $535 per sq m) for dairy, and $70 per sq ft (about $749 per sq m) for frozen foods [40]. Additional costs arise from inventory holding and handling. The inventory holding cost defines a cost factor for each unit of SKU per unit time. Handling costs are comprised of labor cost multiplied with handling time. The handling time significantly depends on the productivity of the personnel. The handling consists of unpacking the cases, disposing the empty packaging, rearranging the products on the shelf, and walking to the storage location of a SKU.
[36]. If the shelf space capacity is smaller than the replenishment lot size, the remaining items have to be returned and stored in the backroom. According to Broekmeulen et al., a store clerk shows output productivity of about 40 cases per hour to 60 cases per hour at an hourly wage of around $5.5. They also state an average sales price per case pack of $15 that results in stacking costs of $0.1375 per case.

II.3 Retail Performance Measures and Drivers

II.3.1 Measures of supply chain performance

Supply chain efficiency derives from high customer service at minimal costs. In this context, a high service level refers explicitly to product availability and the satisfaction of customer needs. The costs associated with delivering high customer service contain all resources that are required to reach certain product availability. Thonemann et al. identify the operational logistics costs and the on-hand inventory as the two main resources. Together with on-shelf product availability, logistics costs and on-hand inventory form the major factors of success [14]. Gilmour presents a group of measures based on a set of capabilities that incorporate a measure for the utilization of technology in logistics processes. In addition, Gilmour addresses the degree to which logistics is accounted for in the overall strategy. The three benchmark groups are process capabilities, information technology capabilities, and organization capabilities [45]. Gilmour states that process capabilities are mainly driven by the customers’ needs and expectations. Therefore, the optimization in cost and performance is directed at increasing flexibility and lowering costs. Beamon divides the supply chain performance measures into two categories – qualitative performance measures and quantitative performance measures [46]. Although the qualitative measures contain performance criteria that are difficult to compare among supply chains, the qualitative measures can be summarized into costs, inventory, and service level. In addition, Beamon adds flexibility as an additional measure and groups the performance measure types into resource, output, and flexibility [47]. Resource is a collective term for measures of inventory level, personnel requirements, equipment utilization, energy usage, and cost. Output is a measure that includes customer responsiveness and the quantity of the final product [47]. Whereas resource and output measures correspond with operational objectives, flexibility measures potential behavior. Beamon describes four flexibility types: volume flexibility, delivery flexibility, mix flexibility, and new product flexibility. In order to assign a responsible management level to each metric,
Gunasekaran et al. group their performance metrics into strategic, tactical, and operational and distinguish between financial and non-financial for all metrics [48]. Gunasekaran et al. assign these metrics to the basic links of a supply chain that are grouped to performance levels. The performance levels are: plan performance, source performance, production performance, delivery performance, and customer service and satisfaction.

Lambert et al. criticize many of the current supply chain metrics claiming that the majority are only internal logistics operations and financial measures (inventory turns and overall profitability). According to Lambert et al., the metrics do not measure performance on key business processes and the effectiveness of meeting customers’ needs. [49]. The measurement of inventory turns fails to capture the key difference in product cost, form, and risk within a supply chain. Therefore, Lambert et al. claim that if opportunity cost of money and the inventory turns are similar, inventory cost at the retail level increases. Therefore, the improvement on inventory turns at the retail level affects the overall supply chain performance more significantly than improvements at another level of the supply chain [49]. They suggest that the measurement of the total inventory carrying costs represents a better performance measure than individual inventory turns because it considers the cash value of the inventory at various positions in the supply chain as well the varying opportunity costs. Supply chain performance measurements that include only internal logistics metrics result in a misalignment with supply chain strategy, e.g. a company may have the strategic goal to provide high customer service but does not measure the service level from a customer’s perspective. Therefore, the ability of supply chain performance to affect customer value is unclear.

According to Lambert et al., a company must analyze whether costs associated with process improvement increase profitability by gaining a greater share of customer business or increasing the supply chain competitiveness. Therefore, they recommend maximizing profitability at each link. This effort will maximize the performance of the entire supply chain [49]. The highest level of competitive attainment leads to higher levels of partner economic performance, customer satisfaction, customer loyalty, and relationship effectiveness [50]. Despite some different perspectives on aligning metrics with supply chain strategy, all metrics for supply chain performance measurement include costs and customer service level as most critical measures.

II.3.2 Measures of retail store performance

Thonemann et al. conducted a study with 28 European retailers to evaluate on-shelf product availability, logistics costs, and on-hand inventory. Their study shows
that the top five companies provide an on-shelf product availability of 98.7%, which is 2.8% above average (95.9%), and 6.8% better than the bottom five [14]. The logistics costs are compared to the annual revenue. The top five spend 2.9% of the annual revenue on logistics costs, which is 1.4% less than average (4.3%), and 2.6% less than the bottom five (5.5%). The on-hand inventory is measured in days and the top five provide inventory for 11.8 days, which is less than half of the average (25.2 days), and a month less than the bottom five (41.6 days). Therefore, the top five European retailers manage to provide a high service level at minimal inventory costs.

The replenishment process represents a major logistical operation within the retail store. Only two processes create value to the customer – the movement of products to the shelf and actual replenishment of the shelf. However, 70% of the time spent on replenishment is used for non-value creating activities such as looking for the right product, looking for a location to stack remaining products, finding a means of transportation (container), and going back and forth on non-standard routes between the backroom and the shelf [14]. As a consequence, the replenishment process is not cost efficient and shelves remain empty longer than required.

II.3.3 Information technology as a supply chain enabler

Most retail supply chains are composed of suppliers, producers, distributors, retailers, and customers interconnected by material, financial, informational, and decisional flows. Each unit within the supply chain aims at optimizing its operational processes according to its individual strategy. However, local optimization at unit level does not necessarily result in a globally optimized supply chain. Even if a unit operates at a local optimum, it may still disregard benefits that could derive from an optimization of the entire supply chain. Along with organizational infrastructure, strategic alliances, and human resource management, Marine views technology as one of the four supply chain enablers [51]. The term technology is not limited to information technology. Technology also comprises materials-management technology for material design, operations, and handling.

The retail supply chain needs to react efficiently to the dynamics of customer demands. Along a supply chain, individual partners have to make supply decisions under uncertainty as the degree of uncertainty increases with the degree of information asymmetry. Information is asymmetrically distributed if the degree of available information varies among the members of a supply chain. Operational and strategic decisions made under uncertainty may differ from decisions that are made with complete and accurate information. Information sharing among supply chain partners reduces the uncertainties of supply and demand for the benefit of all
organizations [52], [53].

Asymmetric information is one of the main causes for the bullwhip effect [54]. The bullwhip effect is a phenomenon that occurs in forecast-driven distribution channels. Because demand forecasts are subject to uncertainty, companies try to compensate for this uncertainty by holding safety stocks. The upstream company experiences a higher fluctuation in demand than the next downstream company, and therefore, has a greater need for safety stocks. If demand increases, the downstream company increases its orders and decreases them again with regressing demand. The bullwhip effect describes the amplifications of variations upwards the supply chain. A demand-driven supply chain that reacts to actual customer demand indicates an improvement, but requires the sharing of demand information along the supply chain. However, the impact of information sharing on the performance of the supply chain depends on supply and demand volatility along with the structure of the supply chain.

Li et al. identifies three levels of information sharing between organizations [55]: The lowest level of information sharing contains the exchange of transactional information such as order quantities, prices, sales, product specifications, quality, and delivery specifications where information technology automates routine transactions between specific buyers and sellers. Operational data such as inventory levels, costs and schedules, production and transportation capacities, lead times, and shipment is shared on the second level and only between two adjacent players – the buyer and the seller. On the highest level of information sharing, strategic information containing point-of-sale information, real-time demand, understanding of market trends, customer preferences, etc. is exchanged while one organization possesses the proprietary information and offers the information to participants to generate strategic benefit (Figure 10).
Lee et al. also specify shared information according to different types: inventory levels, sales data, order status for tracking and tracing, production and delivery schedule, and other information sharing [53]. Lee et al. claim that information about inventory levels is the most common shared data between supply chain partners. This commonly shared element among supply chain partners, including retailers and manufacturers, derives from the risk associated with optimizing their inventory independently. Attempts to independently manage inventory levels risk carrying higher safety stocks than required or possible stock-outs due to miscalculation. In order to optimize its production schedule, the upstream company requires access to the inventory levels at the downstream company. In return, the downstream company benefits through increased service levels at lower overall inventory [53]. Sharing inventory information contributes to lowering the total inventory level along the supply chain and facilitates a reduction in costs.

In traditional supplier-buyer relationships demand information is exclusively communicated in the form of orders. However, order information is usually pre-processed information from the retailer and may be larger than the actual demand. Inconclusive demand information affects the decision-making of the supplier. For example, distortion in demand information may cause suppliers to mismanage production volumes, leading to supply chain inefficiencies. Alternatively, if POS data is shared with wholesalers and manufacturers, they can plan better and adjust their production schedules and inventory levels. In addition, downstream companies
are usually in a better position to forecast future demand. When downstream companies share this information with upstream companies, the production planning process is simplified.

In order for a buyer to compensate for a lack of shipping information with regard to arrival time and quantity of goods, the buyer must order early or place additional orders. Both methods may result in additional costs for inventory holding and shipping, thereby reducing the consumers' value proposition. Consequently, the sharing of shipping information may improve customer service by providing high product availability at minimal cost.

Stefansson defines a logistics information system as follows [56]:

*A logistics information system is an interacting structure of people, equipment, and procedures which together make relevant information available to the logistics manager for the purpose of planning, implementation, and control.*

In order to share information across a supply chain, an information system architecture is required that allows linking information systems of supply chain members through a priori agreed on information interfaces (standards and communication technology). Such information system architecture extends the use of information technology from supporting internal operations to enabling collaboration among supply chain members.

Electronic data interchange (EDI) is a technology that was developed to exchange common types of data between companies based on the “quick response” (QR) initiative by general merchandise retailers and their suppliers [16]. In 1992, the initiative for efficient consumer response (ECR) produced a set of best practices to further improve overall supply chain performance with little changes to EDI systems [16]. ECR encompasses continuous replenishment (CRP), direct store delivery, category management, activity-based accounting, integrated electronic data interchange, and computer-assisted ordering [57]. First developed in the 1980s, EDI has been used to automate routine operations such as posting orders, invoices, shipment notifications, backorder status, etc. The data is transmitted over electronic media such as the internet. A significant reduction in delivery time in comparison to traditional data exchange by mail or fax is the main advantage of using EDI. Additionally, electronic data may be processed automatically upon reception. However, efficient use of EDI communication is only possible if the communication channels and the data transmission are reliable and the integration seamless. The investments that are required to implement EDI communication technology and the costs for installation and maintenance of value-added networks (VANs) are so significant that small and medium sized companies may be excluded from
participating.

Vendor managed inventory (VMI) is a technique designed to improve supply chain performance. According to this technique, the vendor (supplier) manages the inventory levels at the retail store and independently schedules deliveries and shipping quantities to prevent stock-out situations. Using EDI, the retailer shares information about inventory and point-of-sales data with its supplier. The supplier usually proposes a certain delivery quantity prior to its shipment for the retailer to approve.

However, Thonemann et al. state that few retailers have adopted this technique because initial trials resulted in deflating results [14]. Their findings show that VMI is only used for 4.5% of the sales volume (on average) (idem). Hugos et al. also point out that salespeople and distributors are skeptical about overall increase in supply chain performance [15].

Collaborative Planning, Forecasting, and Replenishment (CPFR) is a more recent technique where suppliers and retailers work out the sales planning together. Based on common forecasts, supplier and retailer plan production, delivery, inventory levels, sales promotions, and marketing. However, companies have incorporated CPFR for only 2.9% of the sales volume, an even smaller degree than VMI [14]. This may be due to significant changes required to comply with CPFR. These changes involve existing operations, a distinct focus on one selected retailer, and large investments in specialized information technology.

The systems to process, manipulate, and display data shared between supply chain partners are selected according to the business operations they are designed to support. The enterprise resource planning (ERP) system monitors orders, production schedules, material purchases, and inventory levels within a company. Procurement systems support the procurement activities between a company and a supplier by providing electronic information about the product assortment, part numbers, prices, etc. Advanced planning and scheduling (APS) systems produce schedules for manufacturing processes in different plants based on plant capacity, material availability and customer demand. Transportation planning systems calculate the optimal distribution of materials. Demand planning systems support companies with their demand forecasts based on algorithms that make use of historical data, promotions and other events that affect customer demand. Customer relationship management (CRM) systems automate customer service related tasks, support the acquisition of new customers, and provide quickly accessible buying patterns of customers. Sales force automation (SFA) systems
allow the management and coordination of sales force activities [15]. Combinations of these systems can form a supply chain management system that contains advanced planning and scheduling, transportation planning, demand planning, and inventory planning applications.

Retailers rely heavily on inventory management systems to track historical demand patterns, monitor inventory levels, calculate order quantities, and determine the optimal safety stock level that balances product availability and inventory holding costs. In order to collaborate more efficiently with the upstream partners in a supply chain, retailers need to achieve higher integration by addressing four technological dimensions proposed by Lee et al. [58]. The dimensions are: information integration (allows sharing information among companies in a supply chain), planning synchronization (supports companies in commonly forecasting future demands and scheduling deliveries), workflow coordination (automates ongoing business activities among supply chain partners), and new business models (may emerge from supply chain integration).

Information sharing among collaborative supply chain partners leads to an increased flow of information, reduced uncertainty and a higher profitability of the global supply chain. This results in a higher quality, cost-effective product for the ultimate customer in a shorter amount of time.

A study conducted by GMA and IBM found that consumer product companies invest about 2.1% of their sales revenue on information technology [59]. This represents a small share of the total budget in comparison to other industries. However, eighty-three percent of business executives view information technology as a strategic asset or a return-providing investment. Sixty-nine percent of the total budget allocated for information technology is spent on running the business and improving compliance (Figure 11). Twenty-one percent are spent on strategic issues with only 10 percent invested in initiatives to generate more revenue [59]. The largest shares of the operating budget are spent on personnel costs (60%) for support and maintenance and software (13%). Forty-three percent of the capital budget used to finance long-term outlays is directed toward software, and 17 percent is used for computing equipment with more than half of the investment dedicated to enterprise computing such as application servers, database servers, etc. [59].

Regarding the reasons to invest in information technology, 86% of business executives cite increases in internal efficiency or productivity; 70% name a better compliance with customer and regulatory imperatives, and 56% mention greater efficiency in interacting with trading partners (idem).
The various electronic systems in place that offer access to detailed store-related data should simplify keeping track of sold merchandise, managing inventory levels, automating transactions, avoiding out-of-stock situations, electronically transmitting precise replenishment orders, and sharing demand forecast information with the supplier to benefit from efficient delivery due to production synchronization with actual demand. However, several in-depth studies on data accuracy have shown that most retailers consider it difficult to capture and maintain accurate sales and inventory data [60], [61], [62], [14]. It is estimated that inaccurate data causes overall profit losses of about 10% through unnecessary inventory carrying costs and lost sales from out-of-stock situations [62].

Raman et al. analyzed the data accuracy at a leading publicly traded retailer and found that 65% of the inventory records were inaccurate at the store stock keeping unit level (e.g. the recorded inventory in the management system did not match the physical inventory) [60]. Additionally, the absolute difference between system records and physically available inventory was on average 35% of the target inventory level for each item (idem). Even if the system record is accurate and the items are at the store, merchandise may become unavailable due to misplacements. For another leading retailer, 16% of all merchandise was unavailable to consumers
because they were misplaced and could not be found at the expected location within the store [60]. Inaccurate inventory records may lead to actual out-of-stock situations while the inventory system still reports a certain quantity of stocks. The impact of inventory record inaccuracy and misplaced SKUs may be severe for retailers that rely on automatic replenishment because an item that is out-of-stock but reported in-stock may never be reordered. Additionally, the newly generated forecast based on replenishment and sales data may be so low that a retailer decides to drop a particular item that is actually very popular. For misplaced items, retailers are also likely to underestimate future demand due to inaccurate observation of actual demand [60].

In their study of sixty-five retailers in Europe, Thonemann et al. found that data accuracy in inventory systems varies significantly among companies. The best performing retailers show data inaccuracy of 4.7% while the worst performers have inaccurate inventory records for 29.4% of the products in their inventory [14]. Inventory inaccuracy derives from scanning process errors, misplacements, improper handling of returns, theft, damages, fraud, and other process errors [63], [61]. Fisher et al. finds that salespersons operating check-out scanners induce significant data inaccuracies [61]. The process of reading barcodes still requires significant manual intervention to guarantee the line-of-sight required for proper operation. The manual alignment of products prior to barcode reading is the limiting factor for the speed of barcode scanning. For multiple units of the same price with slight variations in type (e.g. flavor), the check-out clerk may scan one item multiple times in order to speed up the check-out process for the customer’s convenience. However, this causes the inventory system to report a lower actual inventory of the one item that was scanned multiple times while the recorded stock quantity for the other(s) item(s) is too high. Most check-out clerks are reluctant to spend extra time for an accurate scanning process because store managers track and evaluate the average rate of scan units but not the accuracy of sales data generated by the check-out clerk (idem).

Another source for data inaccuracies arises from improper handling of returns. When a customer returns an item because s/he has purchased the wrong type (e.g. size, color, etc.) the clerk is supposed to scan both the returned item and the new item. However, the clerk may decide to just hand out the new item without scanning both items in order to increase the speed of the return handling process. This common practice of handling returns results in inaccurate stock quantities for both items.

Product variety and inventory level may influence the data accuracy in inventory systems. As the variety of SKUs at a store increases, it becomes more likely that
multiple items with high similarity are in a store’s assortment. Employees at the
distribution center who prepare the individual deliveries to the stores may
accidentally pick the wrong items without noticing. Also, employees at the retail
store will find it more difficult to identify misplaced items and to detect out-of-stock
situations and discrepancies between actual and recorded inventory if the product
variety is high. This is further amplified by a high floor employee turnover [61].
New store employees are less familiar with a store’s assortment and the
replenishment process that may result in more errors during the replenishment
process and lower receptiveness for missing or misplaced items.

Retail stores are not the only source of data inaccuracy. Fisher et al. reports that
one company suffered data inaccuracy of 29% on all items caused by incorrect
deliveries from the distribution center. The quantities received at the retail store
differed from the replenishment order because warehouse employees were sloppy
when assembling the mix of SKUs [60], [61]. Additionally, Raman et al. found that
deliveries received directly from the manufacturer showed substantially lower error
rates than those received from the company’s distribution center. They argue that
managers are discouraged to carefully check the accuracy of deliveries from the
distribution center because they do not receive credit for items shipped in error that
cost less than a certain amount [62]. However, store managers check for
manufacturers’ errors because they can receive credit for incorrect shipments
(idem).

Some retailers try to improve a store’s data accuracy through periodic inventory
audits. Physical audits are scheduled approximately once or twice a year. These long
cycle times support the extensive accumulation of errors in the inventory system.
However, audits are often performed to measure the shrinkage of inventory and the
associated monetary loss instead of eliminating incorrect inventory records [62]. By
focusing on the value of the total inventory, an audit may be successfully completed
if the manual audit confirms the total value of products as it is recorded in the
inventory system. A successful audit is possible despite the fact that the actual
product mix may differ greatly from the records.

Fisher et al. describes another approach observed at Staples office-supply
superstore that ensures data accuracy of sales and inventory data, thus preventing
unnoticed out-of-stock situations [61]. The “zero balance walk” is a practice where
employees walk through the store each day and visually check for out-of-stock
situations. For each stock-keeping unit that was detected out-of-stock, the
employees generate a stock-out card and a sticker that is placed at the item’s
assigned shelf location. Other employees analyze the events that resulted in out-of-
stock situations. If the cause was found to be incorrect inventory records, the data is adjusted. The “zero balance walk” not only helps detecting out-of-stock situation and improving data accuracy but also provides a means to measure the store’s performance on consumer service level.

II.4 Out-of-stock

The term out-of-stock solely describes a situation in which merchandise is unavailable for sale. The term does not, however, provide information about whether products will become available for sale again (temporarily out-of-stock) or whether they were removed from the product assortment (permanently out-of-stock). Out-of-stocks (OOS) may occur along the entire supply chain but for the scope of this thesis, the descriptive term “out-of-stock” is limited to on-shelf product availability and is defined as the temporary unavailability of an item on a retail shelf at a particular moment in time.

II.4.1 Stock-outs in retail

According to ECR Europe, product availability at the retail store is approximately 90% to 93%, which is about 6% to 9% lower than at the manufacturer level [33]. A study by Andersen Consulting and the Coca-Cola Retailing council found an average out-of-stock rate of 8.2% in a study covering 11 categories in 10 stores in the U.S. during one month [64]. Gruen et al. state that the average out-of-stock rate from 40 studies that reliably reported OOS was 8.3% [35]. However, they also state that the results may be affected by slightly varying measurement methods. Gruen et al. provide an average low of 4.9% and an average high of 12.3%. ECR Europe distinguishes between categories with high demand, which show average OOS rates above 9%, and lower demand categories with average OOS rates below 5%.
Thirteen of the studies covered by Gruen et al. report variations in out-of-stock rates depending on the time of the day and the day of the week. The highest OOS rates are recorded in the evening (after 8:00 p.m.) and the lowest in the early afternoon. Similar day times for highs and lows were found by a study from the Grocery Manufacturers of America [65].

OOS rates vary significantly throughout the week. According to Gruen et al. the OOS rates reach highs on Sunday and Monday and a low on Saturday [35]. They claim that all examined studies show a declining OOS rate from Sunday (and the carry-over to Monday) to Saturday. The high OOS rates for Sunday and Monday are explained by two factors. First, the weekend is the heaviest shopping time. Second, reordering and deliveries do not occur until the following Monday or Tuesday. For countries where stores are closed on Sundays, high OOS rates on Monday are explained by a delay of restocking until Monday. The OOS rates decline over the week because of restocking and preparation for the heavy shopping days on Saturday and Sunday. However, Saturday shows the lowest OOS rates despite being one of the heaviest shopping days. Gruen et al. claim that store managers employ extra labor and make use of backroom inventory to replenish empty shelves on Saturdays, while labor is normally at a lower level on Sundays and on-hand inventory begins to deplete.

The study by ECR Europe shows differing results. According to their study, Friday, Saturday, and Tuesday are the shopping days with highest OOS rates followed by Monday, Wednesday, and Thursday. The significant OOS rates for
Friday and Saturday are explained by the high shopping activities on these days. Friday and Saturday account for 43% of a week’s sale. Low product availability on retail shelves is further explained by the fact that a retail store does not receive deliveries over the weekend. No weekend deliveries results in low on-hand inventory in the backroom. Although deliveries are made in the beginning of the new week, the shelf availability does not increase immediately because the shelf replenishment is time consuming.

Out-of-stock rates vary significantly among product categories. Gruen et al. identified six categories where reliable OOS rates where reported. Hair care products show highest world average OOS rates (9.8%) followed by laundry products (7.7%), diapers (7.0%), feminine hygiene products (6.8%), toilet tissues (6.6%), and salty snacks (5.3%) [35]. These findings suggest that OOS rates depend on the product category. Gruen et al. found variations within product categories as well. Therefore, product categories are further separated into fast-moving consumer goods and slow-moving consumer goods. The fast movers show a significantly higher OOS rate over the other items of 50% to 80% (idem).

Spikes in OOS rates are detected for products on promotion. Gruen et al. state that OOS rates of promoted items versus non-promoted items is approximately 100% higher. They support their claim with findings by ECR France that reported an increase in the OOS rate of 75% over non-promoted items. In addition, they cite the 1996 Coca Cola study and the 2002 GMA DSD study that both reported a doubling of OOS rates.

The duration of an OOS situation is a critical measure because OOS periods greatly impact the service level. Although high service level is important, it is crucial to have those products available when consumers return to the store after postponing a purchase due to unavailability during the previous visit. Twenty percent of all OOS situations do not exceed 8 hours. Twenty-five percent of OOS situations show durations of 8 hours to 1 day. Thirty-six percent fall in between 1 day and 3 days, while 19% of OOS situations exceed 3 days. Hence, only about 20% of out-of-stock items are replenished during the opening hours of a store. With 55% of OOS situations showing duration of more than 1 day, a returning customer is likely to face the same OOS situation again.

II.4.2 Consumer responses

A study conducted by ECR Europe identified product availability to be the third most important issue after shorter queues for check-out and more promotions [33].
ECR Europe reports that product availability is a central concern of consumers. If a customer is confronted with product unavailability, s/he will eventually purchase a substitute product or go to another store that satisfies her/his needs. An American study shows that 40% of the customers chose to substitute an unavailable item with another one (same or different brand), while 32% buy the same item at another store [35]. Only 17% of the customers delay the purchase and come back another time (idem). For Europe, the findings for store switching are similar: 31% of the customers confronted with unavailability for the first time switch stores. If they are confronted with unavailability a second time, 50% of these customers decide to shop at another store. This number increases to 69% for customers that face product unavailability a third time. The option of delaying the purchase was not given as an option in this study [33]. Therefore, the number of customers that chose a substitution is significantly higher – 69% for the first time, 50% for the second time, and 31% for the third time (idem).

Retailers, especially at the end of the week, should seek to avoid OOS because consumers faced with closed stores on Sunday may cancel the purchase or buy at another store instead of postponing the purchase. (Note, the source from the 2003 ECR Europe study only provides numbers for the combination of “no purchase” and “store switching”. However, it can be assumed that the consumer reaction of “no purchase” remains at a constant level, and hence, it can be subtracted from the numbers given in the study).

Verhoef et al. analyzed consumer reactions in eight different product categories and found that the results vary significantly [66], [67]. For example, the majority of consumers that cannot find their preferred choice of cigarettes prefer to try another store instead of switching brands. However, for eggs only 5% would go to another store. Verhoef et al. also found that brand- and product-related antecedents are much more important than store antecedents (idem).

The critical measure for consumer reactions is a combination of opportunity cost (the cost of not being able to consume the product immediately), substitution cost (decreased use of a less-preferred alternative), and transaction cost (time and effort to buy item at another location). If the opportunity and the substitution costs are high but the transaction cost is low, the consumer is likely to buy the item at another store. If the opportunity cost is low and the substitution cost is high, the consumer either delays the purchase in case of low transaction costs or does not purchase the item in case of high transaction costs. The consumer will buy a substitution if the opportunity costs and the transaction costs are high but chooses the same brand or another brand according to the substitution costs (Table 1) [68], [35]. Consequently,
product availability is critical to achieve high service level and good customer loyalty.

Table 1: Consumer reactions according to opportunity, substitution, and transaction costs [35].

<table>
<thead>
<tr>
<th>Opportunity cost</th>
<th>Substitution cost</th>
<th>Transaction cost</th>
<th>Consumer reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Buy at another store</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Delay purchase</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Substitute (same brand)</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Substitute (another brand)</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>No purchase</td>
</tr>
</tbody>
</table>

Consumer reactions to out-of-stocks vary among consumers according to their perception of store loyalty, time constraints, shopping frequency, and shopping attitude. Store loyalty mainly contributes to the psychological substitution costs of store switching [68]. Campo et al. claim that consumers, who do not want to spend a lot of time on shopping or even show a shopping affinity, are less likely to go to another store or return to the same store at a later point in time. The type of shopping trip (major or minor) also affects the store switching consumer reaction. Consumers with an extensive shopping list usually have higher objective substitution costs when they have to buy particular items on their list at another store. If they cannot get a particular item from an extensive shopping list, they usually experience a lower opportunity cost of reduced consumption [68]. Category consumption rate and product importance increase the opportunity costs of a consumption reduction (idem).

Despite the extensive research on consumer reactions to out-of-stocks, the results show some weaknesses. First, these studies focus on certain aspects, and therefore, the generalization of the findings is limited. For example, they focus on a very limited number of product categories instead of all categories. In addition, the studies disregard of the amount of purchases [69]. Second, these studies were conducted based on single occurrences of out-of-stock. The detection of single occurrences does not allow for generalizations because long term consequences cannot be derived (idem).

II.4.3 Economic impact

Out-of-stock situations result in a 3% diminution of sales [35] which leads to a reduction in profit of about 10% [62]. The impact of out-of-stocks on lost sales
varies with product category – the fastest moving 25% of items account for 66% of lost sales [64]. A retailer may lose a potential sale if a customer faces an out-of-stock situation. The potential sales loss occurs because the consumer may decide on not purchasing a substitute (no purchase from this category) or decides on buying the item at another store. If the customer decides on buying a substitute, the retailer is likely to face a partial loss. This results from the fact that customers tend to pick a cheaper product than the one they intended to buy in order to reduce the risk of a mispurchase [35]. Even if the customer decides on postponing the purchase, the retailer may face a loss of sale because the customer may later decide on buying that product at another store. Emmelhainz et al. state that retailers lose about 13% of their customers due to out-of-stocks [70].

The manufacturer faces a direct loss when a customer is confronted with an OOS situation and decides on buying a substitute. The manufacturer faces a complete loss of sales if the customer chooses a substitute from another brand and a partial loss if the customer decides on a smaller and/or lower price substitute of the same brand.

Indirect losses to both retailers and manufactures may occur if the customer switches the store and the brand due to an unsatisfactory service level. For the retailer, store switching results in the sales loss of an entire shopping basket as opposed to just one item. In approximately 45% of OOS occurrences, consumers’ reactions result in negative consequences for the retailer [66].

According to the 2003 ECR Europe study, European consumers are more likely to switch to a substitute (53%) than to switch stores (21%). Gruen et al. finds that 40% of the consumers buy a substitute and 32% switch stores [33], [35]. However, each time a customer faces an out-of-stock situation, s/he is more likely to switch the store [33].

Retailers face the greatest direct loss in the categories of diapers, feminine hygiene products, tooth paste, and shampoo/hair care products. Manufacturer losses are most significant in the categories of salty snacks, paper towels, toilet tissues, and laundry products [35]. Consequently, a retailer should pay significant attention to effectively managing those categories that result in highest potential losses regarding out-of-stock situations. The impact for retailers due to OOS of low-equity brands is small. Therefore, the effort on OOS reduction should be directed towards high equity brands [66]. Unfortunately, the categories that are most critical for a retailer differ from the categories that result in the highest potential loss for the manufacturer. Therefore, when retailers and manufacturers try to coordinate their efforts to avoid out-of-stock situations for specific categories, they do not peruse the same goals.
II.4.4 Root causes

The root causes for OOS are divided into four categories according to the root location of the cause – store related (72%), distribution center (10%), manufacturer or retail headquarter (14%), and others (4%) [35]. Store related causes are subdivided into store ordering (34%), store shelving (25%), and store forecasting (13%) (idem).

ECR Europe provides a more detailed analysis of root causes and lists 13 major causes – delisting by staff, inventory inaccuracy, damages and shrinkage, shelf replenishment, delivery schedule, operations at the distribution center, supplier reliability, store ordering, distribution center ordering, incorrect master data, forecasting inaccuracy, other distribution center/supplier related causes, and other supplier related causes. However, ECR Europe also concludes that a majority of the problems are directly related to retail operations. These causes are store ordering (35%), delisting by store staff (30%), shelf replenishment (12%), inventory inaccuracy (11%), and other causes (12%). In comparison to Gruen et al., ECR Europe lists “delisting by store staff” and “inventory inaccuracy” as separate causes. While “inventory inaccuracy” contributes to both problems “store ordering” and “shelf replenishment”, the delisting of products by store staff is a separate category, which represents a significant cause for OOS. Gruen et al. have not identified “delisting by store staff” explicitly but according to the definitions of their categories, “delisting by store staff” is understood as a planning related problem that contributes to both retail store and distribution center activities.

About one fourth of the causes for OOS are due to poor replenishment practices. Although the products are in the store – in the backroom or in another area of the retail store – the consumer cannot find the item at the expected location because that particular shelf was not restocked. These situations result in a 25% diminution of profit [62]. The reasons why store staff fail to timely replenish shelves are numerous – busy staff, poor backroom operations, lack of shelf-edge labels, no or insufficient out-of-stock checks, plan-o-gram issues, and shrinkage (damaged products and theft). Therefore, it is important to pay attention to the stock levels on the shelves and to manage the store operations efficiently in order to provide a high service level.

Almost half of the causes for OOS are due to ordering and forecasting problems that result in insufficient stocks at the retail store. Surprisingly, carrying higher inventory levels in the backroom does not reduce OOS rates. On the contrary, it was found that stores with higher on-hand safety stocks show higher OOS rates [33], [35]. It is assumed that higher inventory levels in the backroom indicate poor replenishment strategies that also affect the shelf replenishment from the backroom.
Figure 13: The root causes of shelf unavailability according to the findings in [35].

An attempt to reduce out-of-stocks requires modifications to current processes – at the store, in the supply chain, and at the manufacturer. These processes and their execution can be supported by technology. The use of technology would increase the visibility of inventory levels and improve the accuracy of the data records. A fast detection of low stock quantities in accordance with efficient execution of shelf replenishment could significantly reduce OOS causes at the retail store. As shown above, the manufacturer’s profit is also significantly affected by out-of-stocks at the retail store. A manufacturer who already invests in a good shelf space location (cf. II.2.1) may invest in technology to provide a high service level and maximize profit.
An efficient replenishment process relies on accurate and timely information about shelf stocks. The automatic monitoring of product availability provides better information at lower cost than manual stock inspection. This chapter examines different approaches to continuous shelf inventory monitoring, develops technology specifically designed for this application, and discusses the data quality derived from these technologies.
III.1 Approaches to Shelf Inventory Monitoring

III.1.1 Automatic process control

The adoption of the barcode technology and the development of standards such as UPC and EAN allowed for electronic identification of product types. Consequently, retailers have gained access to a significant amount of data such as sales data, inventory levels, order quantities, deliveries, lead times and shipments, and production schedules – information that is partly shared through and with supply chain partners. The point-of-sales (POS) system at the retailer’s check-out generates data on sales that will be used with store delivery data to estimate the on-hand inventory at the retail store. Subtracting the recorded sales data from the delivery data and adding the result to the current inventory level should allow for an accurate estimation of the on-hand inventory level. However, the estimation may be distorted by theft and misplacements. (Misplacements lower product availability because they cannot be located and found at the retail store. Therefore, the items are treated as unavailable inventory even though they do not reduce the overall inventory level).

Although the utilization of POS data should simplify inventory and ordering management, the analysis on data accuracy in the previous chapter (cf. II.3.3) has revealed that information data at retail stores is highly inaccurate. As stated, inaccurate inventory records may result in actual out-of-stock situation that remain unnoticed by the inventory system. Automatic replenishment systems that rely on accurate inventory records may never trigger a replenishment process despite an actual out-of-stock situation because the incorrect system record shows an inventory level greater zero.

In addition, some retailers attempt to improve the accuracy of inventory records through manual audits. Physical audits are cost intensive and require significant labor. Consequently, cycle times for physical audits, often performed once or twice a year, are usually low. Moreover, this process is susceptible to human errors. In between audits, errors in the inventory system accumulate and may lead to several out-of-stock situations due to incorrect inventory data. Therefore, it is desirable to significantly reduce the cycle time for audits while keeping the process costs low. This goal may be achieved through the automation of the auditing process. Such automation would increase the speed of information gathering, reduce labor cost, and limit the human errors introduced to the system during an audit.
III.1.2 Information granularity

As mentioned in II.4.4, one of the major root causes for OOS situations is poor replenishment practices. Because the store’s inventory management system carries an overall inventory that combines backroom and sales floor inventory, the sales clerks cannot rely on the electronic inventory system to alert them to a low stock on the retail shelf. The inventory management system fails to separate between the two inventories (Figure 14). As a result, it was found that many items that run out-of-stock on a retail shelf can still be found at another location in the retail store, e.g. in the backroom.

In order to enhance the inventory management system with accurate backroom and sales floor inventory levels, the movement of merchandise from the backroom onto the sales floor could be recorded. However, the scanning process of all merchandise or shipping containers requires significant manual intervention. Because of such manual intervention, the process is slow and expensive. Currently, barcode scanning of merchandise is limited to record deliveries and sales. Therefore, this data does not allow for a separation of backroom and sales floor inventory.

Academic research has made significant progress in the development of enabling technologies to automatically distinguish and identify objects. These technologies retrieve information on an object’s location and context. Additional sensors, improved networking capabilities and ever smaller and more powerful distributed platforms lead to faster and more accurate information. This technology offers potential for many business applications in the retail industry where detailed measurements are a precondition to successful management.

Currently, leading retailers (e.g. Wal-Mart, Target, Tesco, and Metro) introduce radio frequency identification (RFID) into their supply chains at case and pallet level for automatic product identification and tracking. In comparison to the barcode, which is limited to the detection of an item’s product type, RFID allows for the identification of individual items. RFID in supply chains makes use of magnetic induction for operation at high frequency (HF) and of emission of electro-magnetic waves at ultra high frequency (UHF). Neither type of RFID requires line-of-sight between the reader and the tag for data transmission. This represents a significant advantage over the barcode technology. The RFID system continuously reads RFID tags in close vicinity and automatically updates the inventory levels. This technology significantly improves the inventory data accuracy in retail inventory systems. Valuable information for inventory management is derived from a combination of distinctive identification, local resolution, and time of detection – basically knowing the position of the item spatially and temporally.
RFID readers positioned at the door between the backroom and the sales floor could automatically capture the movement of goods and update the backroom and sales floor inventory accordingly. Figure 14 illustrates the separation between backroom and sales floor. If low stocks are detected on the sales floor, a store manager knows whether the required items are available in the backroom and may demand the sales clerks to replenish a shelf from the backroom. Additionally, the sales floor staff that replenishes shelves according to the response-based logistics policy (cf. II.2.3) also benefits from accurate information on backroom inventory. The staff no longer has to estimate the backroom inventory level when adding items to the picking list because they possess exact knowledge of the available quantities. The increased visibility of inventory levels will make the replenishment process more efficient.

Figure 14: The retail store is separated into two storage areas – backroom and sales floor.

By distinguishing between backroom and sales floor inventory, a retail store manager can better control the stock quantities and anticipate OOS situations before they occur. However, the sales floor inventory level remains highly susceptible to shrinkage and misplacements unaccountable to an inventory system. Reliable detection of empty shelves occurs if the quantity of individual products at a certain location is precisely determined. However, this type of detection requires an increased level of detail on location information (Figure 15). A variety of sensors are feasible to detect the quantity of products on a shelf (cf. III.2 and III.3). These sensors are categorized into object-centered sensors and sensors that are embedded in the environment. Object-centered sensors include sensors such as RFID tags or barcode label that are directly attached to or printed on an item. The detection of an
item occurs indirectly through an intermediate device (the sensor). Assuming a strong bond between the sensor and its dedicated item, the detection of the sensor allows for the assumption that the corresponding item is at the same location at the same time as the sensor. *Environment-embedded* sensors are incorporated into or attached to obstacles in the environment from where they detect items directly (as opposed to object-centered sensors where the detection occurs through an intermediate device). Examples for environment-embedded sensors are cameras, ultrasound transceivers, infrared devices, or pressure sensors. These sensors detect and distinguish items according to their physical characteristics, e.g. form factor, weight, outline, material, color, etc.

This thesis assumes a “smart shelf” approach to on-shelf inventory monitoring where retail shelves are equipped with technology to support determining the stock quantity on display. For object-centered sensors and specifically for RFID, only a part of the monitoring technology is integrated into the shelves such as the RFID antenna and data processing unit while the sensors, the RFID tags, are attached to the items on display. For environment-embedded sensors, the entire technology required to monitor products on display is integrated into the shelves on which the products are displayed.

![Figure 15: The retail store’s inventory system distinguishes between backroom and individual product shelf space.](image)

The investment in technology necessary to reduce out-of-stock situations and increase the customer service level may not solely be covered by retailers. According to Drèze et al., manufacturers already invest in preferred shelf locations
[40] (cf. II.2.1) and may be willing to invest in product availability monitoring technology for retail shelves to assure high availability of their products.

Drèze et al. found that the number of facings does not correlate with a customer’s attention towards a specific product. Therefore, the number of facings may be reduced to a minimal level. The reduction of the allocated shelf space reduces the inventory holding costs, but efficient shelf replenishment must ensure that OOS situations do not occur despite the lower product quantities. Manufacturers may use product availability monitoring data to evaluate a store’s replenishment performance and to determine the root causes for OOS situations more accurately.

The introduction of technology aims to provide the retail store manager with accurate and timely information about OOS situations. The store personnel are directed to replenish specific products instead of spending an inordinate amount of labor time on manually detecting low stocks on shelves. By pinpointing safety stocks to support items that are likely to become out-of-stock, safety stocks become a part of the supply chain flow rather than inventory that is held by the retailer [35].

### III.1.3 Feedback control loop

An automatic control system relieves operators from making operational decisions commonly based on the operator’s perception on how to best operate a process. A monitoring system measures the current value of the process, compares it to a reference value and adjusts the actuator to compensate for any divergence due to disturbances (Figure 16). Hence, a product availability monitoring system determines the on-shelf stock quantity (current value) and compares it to a threshold value (reference value). If the stock quantity has dropped below the threshold value due to consumers taking items from a shelf (disturbance), the store’s personnel (actuator) are informed to replenish that shelf.

Franklin et al. defines control in the context of automatic control systems as:

> [...] the process of causing a system variable to conform to some desired value, called a reference value [71].

In a closed-loop control system, a feedback controller monitors the current value and compensates for disturbances to the system. Therefore, feedback is defined as:

> [...] the process of measuring the controlled variable and using that information to influence the value of the controlled variable [71].

Consequently, a closed-loop control system controls a process through a rational arrangement of process equipment with measuring/analytical devises, controllers, actuators, process control systems, and computers [72] to ensure that the process performs according to market conditions in terms of availability of material, energy,
labor, and customer demand.

The functional block diagram in Figure 16 illustrates the mathematical relationship among the different variables. The goal of feedback control is to use the principle of feedback to cause the output variable of a dynamic (time varying) process to follow a desired reference variable accurately regardless of the reference variable’s path and of any external disturbances or any changes in the dynamics of the process [71].

**Figure 16: A component block diagram of an elementary feedback control loop.**

Controllability and observability are a major concern for each system. A controllable system offers the possibility of forcing the system into a particular state by applying a certain control signal. If, however, the system is uncontrollable, the control signal will never be able to stabilize the system. Observability describes the capability to observe a system’s state by measuring its output. If the system is not observable, the actuator or controller will never be able to adjust the behavior of the closed-loop system.

Despite controllability and observability, a feedback control loop still suffers from a time lag between the observation of a deviation in the controlled variable and the effect of the corrective action. The time lag occurs because the control system does not provide predictive control actions to compensate for the effects of possible disturbances. If large and frequent disturbances occur, a control process may be forced into a continuous transient state – never attaining a steady state.

The control system is also subject to time delays until the corrective actions are observed at the output. Delays in the transportation of goods (items moved from the
backroom to the shelf on the sales floor) and the cycle and processing time of the measurement device constitute the sources for time delays of the control system. Time delays may introduce difficulties into the control process such as delayed detection of disturbances and long periods of time until corrective control actions affect the process variable.

In the following two subchapters, radio frequency identification (RFID) and weight-sensitive foam will be introduced as sensors into the replenishment control system.

III.2 RFID Technology

III.2.1 Introduction to RFID

More than two decades ago, the barcode triggered a revolution in identification systems. Detailed information about the flow of materials or goods through automatic identification increases the visibility and efficiency of existing processes in various industries. Although the barcode labels are extremely inexpensive, they were found to be inadequate for many applications due to their low data storage capability and the fact that they are not reprogrammable [73]. A technically optimal solution is the digital storage of data on a small device attached to the object that needs to be traced. The data on the device is accessible through a wireless communication channel. Because these devices are distinguished through individual identification numbers embedded in the tag and transmitted to readers through radio waves, these systems are called Radio Frequency Identification (RFID) systems.

The recent introduction of RFID transponders into leading retailers’ supply chains has had a tremendous impact on their ability to manage the flow of goods. Tags on cases and pallets increase the supply chain’s visibility and allow for accurate tracking and tracing. RFID systems wirelessly capture data from RFID tags that are attached to objects within close vicinity. The signals are transmitted without the requirement for line-of-sight using an electromagnetic challenge/response exchange method.

A RFID system is comprised of two main components – the tag (or transponder) and the reader (or interrogator). Both components connect to antennae to wirelessly exchange data (the interrogator may be a read or read/write device. (In this thesis; the data capture device is always referred to as the reader). The RFID reader contains a radio frequency transceiver module and a controller unit. Most readers are equipped with an additional serial interface to forward the captured data to
another system for further processing. The tag consists of a tuned resonant circuit and a low-power CMOS integrated circuit (IC). The IC chip contains an analog RF interface, an antenna tuning capacitor, a RF-to-DC rectifier system, digital control, EEPROM or SRAM memory, and a data modulation circuit.

More sophisticated tags have security features implemented such as security gates or a security bit [74]. The functionality of a tag can be further enhanced through additional sensors that capture data from the environment such as temperature, humidity, vibration, etc, and the collected data is transmitted over the RFID communication channel.

RFID tags are separated into three categories – active, semi-active, and passive – according to their power source. Active tags are equipped with a battery that supports long-range communication and allows powering additional sensors on the tag. Passive tags completely rely on an external power source. The power required to operate a passive tag is wirelessly transferred from the reader through coupling. Therefore, the passive tag is only activated when it is within the interrogation zone of a reader. Semi-passive tags are an intermediate form – they make use of a battery to power the on-chip electronics, but rely on the reader signal for communication and data transmission.

Other important differentiation criterions for RFID systems are the operating frequency, the physical coupling method and range of the system [73]. RFID systems operate at a wide range of frequencies from 135KHz to 5.8GHz. Electric, magnetic, and electro-magnetic fields are used for power and data transmission. Systems that operate at low frequencies (LF) of 135KHz and at high frequency (HF) of 13.56MHZ are inductively coupled systems. These systems are referred to as remote coupling systems with read and write ranges of up to approximately 1m [73]. The magnetic near field is an energy storage field. The field strength path of a magnetic antenna along the coil x axis follows the relationship 1/d³ in the near field (damping of 60dB per decade of distance) [73]. Long-range systems describe RFID systems that operate at read ranges significantly above 1m. In contrast to the inductively coupled LF and HF systems, long-range systems use electro-magnetic waves at ultra high frequencies (UHF) of 868MHz, 915MHZ, and 950MHZ (Europe, USA, and Japan, respectively) and at microwave frequencies of 2.5GHz and 5.8GHz. The electro-magnetic far field is an energy propagating field. The field strength decreases according to the relationship 1/d as distance increases (damping of just 20dB per decade of distance) [73]. The range limit for inductively coupled systems and the beginning of the far field is roughly given as λ/2π (idem). The
majority of long-range systems makes use of backscattering (modulated signal reflection) to transmit data. Passive backscatter transponders have typical read ranges of 3m, while active backscatter transponders reach read ranges of 15m [73].

In a general operating setup several RFID transponders may be concurrently present in the interrogation field of the reader. The reader uses a broadcast communication scheme to transmit data to the transponders. The transmitted data is received by all transponders simultaneously. In order to avoid interference during data transmission from the tags to the reader, the individual channel access for transponders is dynamically allocated by the interrogator according to a multi-access scheme. Numerous procedures have been developed to manage the communication channel allocation. Finkenzeller lists four basic procedures – space division multiple access (SDMA), frequency domain multiple access (FDMA), time domain multiple access (TDMA), and code division multiple access (CDMA) [73]. These procedures are designed for uninterrupted communications from and to the tags, and a channel capacity that remains assigned for the duration of the relationship. However, RFID communication is characterized by short periods of data transmission followed by longer and unequally distributed periods of inactivity. This poses a challenge to the basic multi-access procedure derived from satellite and telephone networks. Media access control schemes frequently used in RFID communication protocols are variations of framed ALOHA that are capable of coping with the channel access characteristics typical to RFID [75], [76].

Various standards have been developed (e.g. ISO 14443, ISO 15693, ISO18000, and EPC), while the Electronic Product Code (EPC) is likely to form the basis for a worldwide standard in logistics. The EPC is a 96-bit tag that contains a number called the Global Trade Identification Number (GTIN). Unlike a UPC number, which only provides information specific to a group of products, the GTIN gives each product its own specific identifying number. Therefore, RFID in combination with EPC facilitates automatic identification (Auto-ID) of a distinctive object in the supply chain.

The information value generated through RFID is further increased if shared with supply chain partners. The EPC network architectural framework is designed to make the distributed information repositories available to subscribers by offering an EPC Information Service (EPCIS) access interface. Multiple EPCIS systems exist due to the complex structure of supply chains and their numerous interrelationships. The location of information sources relevant to a certain identity (EPC) is provided by the Discovery Service.
III.2.2 The impact of RFID on the retail sector

With the deployment of RFID at the case level and the combination of RFID data with POS data, the inventory accuracy will improve significantly. Picking lists can be generated based on sales data and accurate backroom inventory that allows concluding the stocks on the sales floor. In order to determine the effect of RFID on out-of-stock levels, Hardgrave et al. conducted a study with 12 test stores that had RFID implemented at the case-level and another 12 control stores that did not use RFID [77]. The test stores were equipped with readers in the backroom and at the doors that lead to the sales floor. This setup allows separating backroom inventory from sales floor inventory. The process of generating a picking list has changed from visually inspecting the shelf stocks to automatic generation based on real-time sales and inventory data.

Hardgrave et al. found a 26% improvement on OOS rates for stores using RFID (from an average weekly OOS of 444 items to 328 items). In addition, they examined the reduction of OOS at the control stores and explained the improvements found at the control store by the altered behavior of people under observation (Hawthorne Effect). The control stores improved by the fact that they scanned their shelves on a daily basis. The improvement of RFID-enabled stores over the control stores was 63%. Hence, the actual improvement of the test stores was 16% (26% · 63%). However, this initial study failed to address several issues that have an effect on out-of-stocks such as lot size and shelf quantity. If the sales velocity (number of units sold per day) of each product is taken into account, a reduction of 30% in out-of-stock was found for products selling between 0.1 and 15 units per day [78].

A study by PricewaterhouseCoopers evaluates the impact of the deployment of RFID on sales improvement and labor reduction [79]. They estimate that an average grocery store could reduce 50% of their out-of-stocks caused by execution issues and about 33% for RFID at case level. The reduction on labor hours is estimated at 100 hours per week for case-level RFID and 200 hours per week for item-level tagging. Store labor accounts for 11.2% of sales, which is improved by 0.5% with case-level tagging and by 0.8% with item-level tagging.

In addition, the study states that the adoption of RFID at case-level has a break-even point at 20% improvement on out-of-stocks and a store labor reduction of 60 hours per week – both well within their estimation on possible improvements. Despite the more significant improvements, which are possible with RFID item-level tagging, the network of reader-system infrastructure is more complex resulting in higher infrastructure costs. Additionally, the number of disposable tags attached
to consumer products will significantly contribute to the system costs.

Jones et al. reports on an extensive trial run by Marks & Spencer where RFID was used to track 3.5 million reusable trays, dollies and cages throughout the store’s refrigerated food supply chain [80]. The company experiences 80% reduction in the time taken to read a stack of multiple trays in comparison to conventional barcode operations. The tags are reusable and the costs can be spread over time. Marks & Spencer estimate that the capital cost of a RFID system will be less than 10% of the annual costs of using barcodes (idem).

Other trials by a clothing manufacturer in the US indicated as much as a 7% increase in sales when RFID was used because of the greater visibility of the inventory on the shop floor [81].

Barua et al. conducted a study that assesses the financial impact of RFID on the retail sector [82]. The overall findings for current adoption levels of RFID at pallet level (9% of sales) and at item level (2% of sales) indicates that sellers could currently derive $12 billion in benefits from existing RFID applications.

The study states that the total reduction on non-supervisory labor costs for the retail industry will be about $102 billion for an adoption of RFID at pallet level of 100%. This reduction would cut the current costs of about $260 billion almost in half. With current adoption levels, the reduction is estimated at approximately $2 billion (about 2%).

According to Barua et al., RFID will also significantly increase inventory turns. Barua et al. state that retail inventory turns fell from 14 in 2002 to fewer than 11 in 2003, suggesting average annual inventory holding costs of $105 billion. They expect that accurate, real-time information and insight into the flow of goods along the supply chain reduces inventory errors, optimizes time and resources, thereby improve inventory forecasting and enhance inventory turns. Wal-Mart expects reduced inventory costs of about $1 billion or 0.4% of sales, and Accenture estimates a reduction in safety stock of about 10-30% [82]. Barua et al. estimate the total reduction in inventory costs that accrues to the retail industry from the use of RFID at pallet level at $16 billion ($1.4 billion at current adoption level), and the reduction in inventory write-offs due to spoilage and obsolescence to $10 billion ($0.2 billion at current adoption level) for RFID item-level tagging.

The effect of RFID on out-of-stock situations and product availability is examined by Barua et al. based on the reduction of human errors, a better match of sales against orders, and more efficient inventory operations due to increased location accuracy of goods. In their trials with RFID pallet-level tagging, Wal-Mart found a
reduction of approximately 16%. Based on these findings, Barua et al. estimate the increase in revenue from reduced out-of-stock situations for the retail industry at $11 billion with corresponding free cash flow savings of $1 billion for an adoption rate of RFID at pallet level of 100%.

AT Kearney estimates a reduction of out-of-stock situations at the retailer of 0.07% of sales for RFID item-level tagging [83], while METRO Group expects a reduction of 9-14% in the number of occurrences [84]. It is also expected that item-level tagging will significantly reduce shrinkage. Retailer’s losses due to product shrink are estimated at 1.75% of sales [85]. Sixty percent of shrinkage is attributed to theft by both employees and customers (idem). According to Smith, 55% of all product theft occurs in the supply chain prior to goods reaching the retail store [74]. Apparently, Metro Group found a reduction in product shrinkage of 11-18%, and therefore, Barua et al. estimate the reduction in retail product shrink for RFID at pallet and item levels at $19 billion. IBM’s estimate on shrinkage reduction aligns with the one made by Chappell et al. of approximately 47% [86], [87]. The reduction in shrinkage will result in higher sales figures and McKinsey estimates an increase of 0.6-1.5% at the high-end apparel retailers [88], which aligns with the estimates by Accenture of 1-2% [89].

Further benefits arise from an inventory reduction and labor cost savings. AT Kearney estimates an inventory reduction of 5% [83], while SAP estimates a reduction of 8-12% [90]. Labor cost savings at retail stores are estimated at 17% [84]. These labor savings are mainly due to reduced manual intervention in inventory audits.

**III.2.3 RFID read rates**

The ultimate goal of applying RFID to retail processes is the increase of the grain of information in order to support decision making. However, information economy is highly dependent on the quality of the data. The attributes of data quality include: accuracy, completeness, timeliness, and consistency [91]. Information quality characteristics can be blurry; and perfect information quality is difficult to achieve (idem). The dimensions of information quality are directly related to organizational information systems [92]. Managers significantly rely on business systems to acquire appropriate information allowing informed decision-making associated with operational and strategic business processes (idem).

The most important criterion for accurate inventory data and hence high information quality is the read rate at which data on merchandise is captured. In
fact, many companies still consider read performance unsatisfactory [93]. Insufficient read rates either result in the distortion of inventory data or require significant manual intervention (e.g., rotation of pallet or repeated passing of reader station) to avoid errors.

GS1 France, a branch of the global organization that designs and implements standards for supply chains has evaluated the general performance of RFID in the frame of precise applications [94]. In their study, they distinguish two main categories – single reading applications and simultaneous reading applications (mass reading). Achieving reliable bulk reading without any disassembling of pallets would provide significant value to the process of data capturing, thereby improving the management of a supply chain.

The read rate performance tests were conducted with a wide range of products. These items included metallic products and packaging, as well as products with liquid content. The tests intended to evaluate the performance on mass reading were carried out in three phases. Each phase included a different stacking of the pallet (homogenous pallet, composite pallet, and heterogeneous pallet). The mass reading was performed utilizing three methods – with a reader portal passed at pedestrian speed and rotary and manual reading at read ranges of 0.2m to 0.5m (Figure 17). The tests were performed at the two frequencies specified in the EPC standard, at HF (13.56MHz) and UHF (860 and 960 MHz).

Figure 17: Mass reading setup (portal, rotary or manual) to evaluate read rate performances [94].

The first tests were conducted with “neutral” products. These products do not interfere with the RFID signal. This “neutral” category includes all products and their packaging that do not contain metals or liquids. The test results for “neutral”
products on homogeneous or heterogeneous pallets show read rates for UHF of 100% while operating on an emission power that complies with the European regulation ETSI. These results on read rates were reliably reproduced.

For products containing liquids, difficulties were reported for detecting RFID tags that are placed in the middle of a pallet and shielded from the interrogator’s signal by the liquids (Figure 18). The tests were conducted at HF and UHF leading to significantly different results. Despite UHF antennae specifically designed to work on signal absorbent materials, maximum read rates of only 70% were achieved. The conditions to reach maximum read rates were rotary reading with careful consideration of tag placement on cases homogeneously stacked on pallets. Both conditions, rotary reading and careful tag placement, represent major constrictions for logistics applications. In addition, it was found that the shape of the product may have an influence on the bulk reading performance. When products do not fill the entire space in the case, and room is left for the RFID signal to penetrate without interference, higher read rates result. Read rate tests performed at HF with products containing liquids showed continuous read rates of 100% (Figure 18). However, the short read range of HF does not allow for a portal read station but requires manual reading.

![Figure 18: Read rate performance on homogeneous pallets stacked with products containing liquids – read rate given for HF and UHF [94].](image)

Metallic products or packaging containing metallic disrupts the proper operation of the RFID antenna and produces interferences. Again, tests were performed at both frequencies, HF and UHF. The results obtained are highly dependent on the product and vary significantly among different types of products.

Products, such as razors or toothbrushes, that contain only small metallic components showed continuous read rates of 100% on homogeneous pallets. Even pallets with car radiators allowed for read rates of 100%, which is explained by the
small material density on a completely stacked pallet. However, for products with large metallic surfaces, read rates for UHF and a portal reader setup dropped to 30%. In contrast to UHF, HF performs well with read rates between 93% and 100%.

![Figure 19: Read rate performance on homogeneous pallets stacked with metallic products – read rate given for HF and UHF [94].](image)

Three additional properties were identified that significantly influence the read performance. The first property is the alignment and organization of products within a logistics unit (e.g. case). The second involves the placement of the reading devices (position and orientation). The third is the alignment of cases stacked on a pallet.

As outlined above, the alignment and organization of products significantly influences the penetration of a case. The tag should be placed in a way that the shielding through the product is minimized. Additionally, best performance of the tag-reader system is achieved when the tag antenna and the reader antenna are aligned on parallel planes. This positioning becomes increasingly complex in cases of heterogeneous pallets where cases with different tag placements are stacked. In this case, the operational performance of the tag is influenced by the products of the case to which it is attached and by the other cases and their tags. Therefore, special consideration must be paid to the general orientation of stacked cases in order to achieve best mass reading performance.

The influence of read rates on inventory records and on the performance of the replenishment process is further discussed in Chapter IV. In that chapter, models for shelf inventory management will be presented that account for inaccuracies in data capturing.
III.2.4 Increased location information

III.2.4.1 Location information
The deployment of RFID to support logistics processes in the retail industry seeks to provide seamless location information of goods. A location system provides two kinds of information – physical and symbolic [95]. Physical location information provides coordinates and altitude data. Symbolic location information encompasses abstract ideas of a location such as “on the sales floor”, “next to the information desk”, “at the delivery dock”, etc. A further distinction is made between absolute and relative location. An absolute location system uses a shared reference grid such as coordinates for all located objects, while in a relative location system an object can have its own frame of reference [95].

Despite the fact that RFID item-level tagging is praised as the technology that provides detailed information about the location of each individual item in a supply chain, the level of local resolution is limited to the places covered by RFID reader infrastructure. The grid of reader deployment determines the granularity of location information retrieved from a RFID tag where the reader’s physical location is assigned to any tag within the range of that interrogator. Consequently, the detection of a RFID tag provides only limited local resolution – in close vicinity to a specific reader. However, the exact physical location of the tag remains unknown. Currently, RFID readers neither resolve the distance nor the angle at which a tag is detected even though this information could be valuable to cluster objects. A retailer who deploys item-level tagging tries to minimize infrastructure costs (grid of interrogators), but still requires detailed information about product quantities on each retail shelf. The retailer may possess accurate product quantity information, but difficulties arise in associating the quantities with a certain shelf. This situation exists because the reader field of an interrogator is usually not directed thus it would cover the largest area possible. However, non-directional RFID antennae limit the local resolution.

III.2.4.2 Related work
Electromagnetic sensing at radio frequency offers capabilities to determine distance and direction of an object in relation to a reference point. The major techniques to derive these properties are time-of-arrival, time difference of arrival, angle-of-arrival, and received signal strength [96]. Based on the distance measurement, the location is estimated through triangulation. Triangulation is accomplished either through lateration that uses multiple distance measurements between reference points and the object or angulation that measures angle or bearing relative to reference points [95]. Research on location systems based on radio
frequency is highly advanced [97], [98], [95], [99]. Other research has focused on mapping and localization through RFID tags that are distributed in the environment [100], [101]. In their research, mobile RFID readers are mounted on robots that navigate according to the local information retrieved through the detection of embedded tags.

The techniques in these papers on RF distance-sensing rely on systems that operate on the emission of electro-magnetic waves such as RFID at ultra high frequency (UHF). In contrast, HF RFID systems operate on magnetic induction [73]. The physical principles of operation at high frequency require a different approach to resolve a tag’s distance from a reader, because current techniques for distance-sensing cannot be applied.

### III.2.4.3 Sensing principles

In inductively coupled systems, the reader antenna generates a strong, high frequency electromagnetic field in order to power the RFID tag and to serve as medium for data transmission. The electromagnetic field penetrates the cross-section of the coil area and a certain part of the emitted field penetrates the antenna coil of the tag in close vicinity [73]. A voltage $u_{\text{tag}}$ is generated in the tag’s antenna by inductance. The inductively coupled system can be treated as a transformer-type coupling with a primary and a secondary coil as long as the tag remains in the near field, which is approximately $3.5\text{m} (0.16 \cdot \lambda_{13.56\text{MHz}})$ from the antenna (idem). A tag within the interrogation field draws energy from the magnetic field generated by the reader. This event results in a voltage change at the reader’s antenna due to mutual inductance. A tag transmits data by switching a load resistor on and off. The switching influences the transformed transponder impedance and causes an amplitude modulation of the voltage $u_{\text{reader}}$ at the reader’s antenna coil. The intensity of the mutual inductance depends on antenna parameters and the distance between two antennae [73]. The intensity of the coupling decreases with increasing distance resulting in reduced signal strength. Consequently, the analysis of the signal strength allows concluding the distance between a tag and a reader.

The assumptions made above hold true as long as the magnetic field can be considered homogeneous. However, this is the case only if the reader’s antenna and the tag’s antenna are in parallel with a common central axis. It is assumed that the dimensions of the tag’s antenna are small in comparison to the reader’s antenna and that displacements away from the common central axis are not extensive. However, if the tag antenna is tilted away from the central axis by the angle $\phi$, the induced voltage is smaller and follows the relationship given in Eq. (1) ($u_0$ is the voltage that is induced when the coil is perpendicular to the magnetic field and $u_{\phi}$ is the actual induced voltage).
Consequently, in order to apply proximity sensing to high frequency RFID systems information about the tag’s orientation in relation to the reader antenna is required. In addition, information about the dimensions of the tag and reader antennae and the number of windings is necessary. While the reader antenna and the tags used in this application remain unaltered after a first initialization of the system, tilt angle and signal strength are subject to change. The orientation of a tag is derived from a tilt sensor or gyroscope directly attached to the RFID tag. Transmitted along with the tag’s ID, the reader computes the deflection with respect to its own orientation. In combination with the signal strength of the received data, the reader estimates the distance of the tag from the reader antenna.

III.2.4.4 System design and implementation

The RFID system includes a tag, a reader, and a data processing unit (e.g. a notebook) which is attached to the reader. Current passive tags do not provide interfaces to attach additional devices. Therefore, a passive RFID tag is built with discrete components to provide an interface that allows attaching a tilt sensor. The reader and the reader antenna are commercially available. The tag consists of a coil to pick up the reader signal, an analog front-end to demodulate and modulate the RFID signals, and a logic unit to interpret the demodulated data and to encode and send tag data. Additionally, the tag contains a tilt sensor to provide information about the deflection of the tag away from the parallel plane of the reader antenna as well as an external power supply, e.g. a battery. The external power supply is required because a tag built with discrete components draws considerably more energy than a regular tag. The tag’s power requirements exceed the energy provided by the reader through the electromagnetic field.

III.2.4.4.1 Analog front-end to de-/modulate RFID signals

The RFID tag implements the widely used ICODE1 communication protocol originally developed by Philips. The ICODE1 protocol operates at 13.56MHz. It is a reader-talks-first protocol that uses amplitude modulation and pulse position coding on the reader-to-tag communication channel. The protocol uses a modulation index of 10% and, in standard mode, data is transmitted according the ‘1 out of 256’ pulse position scheme. The value of the transmitted byte is encoded in the position of the pulse. In a set of 256 consecutive positions within a pulse, the value of the signal at one position differs from all the others. This position represents the value of the byte (Figure 20). The data transmission is preceded by a start pulse of 9.44μs that signals the demodulator of the RFID tag that a new sequence of data is sent. The demodulator on the tag determines the bit position by measuring the elapsed time.
between the start pulse and the next detected pulse. Each bit has a length of 18.88μs. The division of the measured time by the sequence of one bit results in the exact bit position and allows decoding the value of the byte. The transmission of a single byte (excluding the start bit) requires 4.833ms. Hence, the transmission of an entire RFID reader command, that consists of 8 consecutive bytes, takes 38.7ms.

![Diagram](image)

**Figure 20: The value of a transmitted byte is encoded in the 256 possible consecutive positions for a pulse [102].**

The analog front-end on the tag consists of an envelope detector to identify relative changes in amplitude, which is a representation for modulated bits. However, amplitude variations can be caused by changes in the coupling between a reader and a tag coil (e.g. tilting of the tag or a change in the distance to the reader). In order to distinguish variations that are due to data transmission from ones caused by changes in coupling, the output of the envelope detector is compared to a relative reference voltage. The reference voltage reflects variations in distance and orientation, but its time constant is chosen so that the voltage remains unaffected by short pulses. Both the envelope signal and the reference voltage are extracted by two RC filter elements (Figure 22). The RC elements are designed to meet the specifications of the ICODE1 air interface protocol with a 10% amplitude modulation and a signal drop of 9.44μs for the modulated signal. In addition, the RC elements account for the negative peak clipping that is caused by the half-wave rectification. The two signals are compared by an operational amplifier element that generates a binary output based on the detection of a modulated signal. The output goes low as long as the signal is modulated and remains high otherwise [103].

In order to transmit data from the tag to the reader, the system makes use of load modulation as described above. A CMOS switch, controlled by the microcontroller, shorts the antenna coil according to a pattern that represents the encoded data.
III.2.4.4.2 Logic unit and signal processing

A PIC16F88 microcontroller that runs internally at 5MHz forms the core of the RFID tag. The clock is not obtained from the carrier signal of the RFID reader but from an external oscillator used to avoid building a clock extractor. The PIC16F88 microcontroller offers two different methods to listen for changes on the comparator’s output – polling and interrupts. While polling proved inadequate because it does not meet the timing constraints for the detection of two consecutive pulses, the interrupt-based approach performs well. The asynchronous interrupt is triggered by a falling edge at the comparator’s output. When an interrupt occurs, the interrupt dispatcher reads the value of the timer that started after the detection of the start bit. Measuring the elapsed time until an interrupt occurs, allows determining the value of the byte transmitted by the reader. After the detection of 8 consecutive bytes, the main routine in the microcontroller code starts decoding the complete command.

The most restrictive time constraints occur when a pulse is sent in the first slot. This pulse follows 18.88μs after the falling edge of the start pulse. The detection of each pulse triggers an interrupt used to measure the elapsed time between consecutive pulses. However, the processing of an interrupt must be completed before the next pulse can be detected. This sets an upper time limit for interrupt processing of 18.88μs. The optimized interrupt dispatcher requires 77 clock cycles that corresponds to 15.4μs. Hence, the design fulfills even the most restrictive time constraints.

In response to a “selected read” or “unselected read” reader command, the tag responds with the data stored at the location specified in the reader command. When sending data from the tag to the reader, the shorting of the tag’s antenna generates only small variations of the voltage at the reader antenna. Therefore, a sub-carrier of 423.75kHz is used to create modulation sidebands at the reader’s antenna for easy detection [73]. Based on this sub-carrier, the tag data is Manchester encoded (idem). The transmission of a single bit requires 37.76μs and an entire byte requires 302.08μs. The data is sent in a random time slot, while the length of the time slot depends on the number of data packets to be sent. The transmission of one data packet requires 1208.32μs.

III.2.4.4.3 Tilt sensor

A dual-axis ADXL202E accelerometer is used to measure gravity. This device allows the determination of the tilt angle in one plane (Figure 21). For reasons of simplicity, the design is limited to the detection of deflections in the y-z-plane. By attaching an additional tilt sensor or by using a gyroscope, the deflection in all three dimensions can be captured.
The analog outputs of the tilt sensor are connected to the A/D converter of the PIC microcontroller that processes and transmits the data. The sensor values show a non-linear behavior of tilt angle and output value. On average, a sensor value unit corresponds to 1.0° to 1.8°. The standard deviation for a fixed tilt sensor is 2.8 units, or 3° to 5°. Due to the properties of the cosine function, small deflections away from the parallel plane do not result in a significant change of the actual inducted voltage.

**Figure 22: Schematic of the tag with the analog front-end, logic unit, tilt sensor, and CMOS switch for data transmission.**
with a resolution of 12 bits. The amplitude of the signal strength is extracted with a simple peak extraction algorithm. In order to cope with signal strength fluctuations, a set of 80 data values is collected at once within a measurement period of about 60s to compute the average amplitude of the signal strength. Additionally, a threshold-based noise filter is used to improve the data quality.

At the next step, the tilt angle $\phi$ that was transmitted to the RFID reader over the RFID communication channel is extracted from the received data. The reduction in induced voltage with respect to the tilt angle according to Eq. (1) is computed. At the last step, the absolute distance is retrieved by comparing the transformed signal strength to a previously established look-up table. The look-up table was created with a setup of a tag in parallel to the reader and a steadily increasing distance. The signal strengths were recorded and stored in a look-up table along with their associated distances with increments of 1cm.

### III.2.4.6 Evaluation and discussion

In order to test the design, a total of 900 samples of the signal strength’s amplitude were collected. The measurements were conducted with an ordinary tag in parallel to the antenna at varying distances. The results are plotted in Figure 23 and show a nearly linear relation between the amplitude of the signal strength and the distance between the tag and reader antennae. Figure 23 also shows the standard deviation for each measurement. The nearly linear behavior is approximated with the linear function described in Eq. (2).

$$y = 4172.8 - 49.79 \cdot x$$  \hspace{1cm} Eq. (2)
Figure 23: The measurements show a nearly linear relation between the extracted peaks $u_0$ and the distance from the reader antenna.

The design was also tested with setups where the tag’s orientation shows a deflection away from the parallel plane (a tilt angle greater zero). For all measurements taken, the transformation according to Eq. (1) resulted in a distance resolution that was within the standard deviation retrieved for measurements performed without any deflections. For example, the design was tested with a tag positioned at a distance of 40cm from the reader antenna and a tilt angle of 30 degrees. The signal strength was measured at 3198mV, which transforms into 3693mV at a tilt angle of 30 degrees (Eq. (1)). This is within the standard deviation for measurements at 40cm of 3672.04mV + 28.10mV (equals 3700.14mV).

The distance measurement procedure allows accurately resolving distances between 33cm and 50cm. Distances outside of this range are not reliably detected. The limitation of operation to a certain range is due to timing imprecision during bit encoding. The Manchester encoding of data generates an oscillating signal at 423.75kHz for one half of the bit length while the other half remains zero (Figure 24). Either the first or the second half shows oscillation depending on whether the bit value represents a logic “one” or logic “zero”. According to the ICODE1 protocol, the length of each bit is 37.76μs, and therefore, the signal oscillates for 18.88μs (half of the bit length) at 423.75kHz. Within these 18.88μs, the signal changes 16 times between “low” and “high” and remains at each value for 1.18μs (Figure 24).
The microcontroller has an instruction cycle time of 0.2μs. Therefore, it cannot generate the required oscillation frequency exactly. The best approximation is 1.2μs (6 instruction cycles), which results in an oscillation frequency of 416.7kHz. Consequently, the elapsed time after 8 oscillations is 19.2μs instead of 18.88μs. To sufficiently approximate the total bit length, the non-oscillating part is shortened to 18.6μs. This results in a total bit length of 37.8μs, which is 0.04μs longer than the exact value of 37.76μs.

The difference in frequency proved not to be critical because the reader’s band-pass filter shows sufficient bandwidth to successfully filter the approximated sub-carrier. The longer bit lengths are compensated for by shortening the transmission of an entire byte by one instruction. This reduces the time divergence from 0.32μs to 0.12μs. With these adjustments the tag successfully transmits data to the reader.

Nevertheless, the distance of operation for data transmission from the tag to the reader remains limited to a certain range due to timing imprecision at the transition point from an oscillating output to a zero output. The timing inaccuracies lead to signal distortions at the RFID reader. Figure 24 illustrates the spikes that occur because of these inaccuracies. As long as the signal strength is strong, the spikes are small in relation to the maximum amplitude and the data is demodulated correctly. With decreasing signal strength the spikes become more significant as their amplitude remains constant, which results in incorrect detections with the presented peak extraction algorithm.

Figure 24: Manchester coding with sub-carrier that represent a 1-bit and a 0-bit [76].
Figure 25: The plot of the scope values for $u_0$ at a distance of 43cm reveals the spikes (circles) occurring due to timing imprecision during Manchester encoding.

Another restriction to this design is caused by the diverging lines of the magnetic flux. The magnetic field is considered homogenous only around the coil axis. Therefore, this design cannot accurately resolve distances for tags that show larger displacements away from the coil axis. However, the assumption of a large coil diameter and an arrangement of the merchandise close to the coil axis are appropriate.

III.2.4.7 Concluding remarks

The ability to resolve a tag’s distance from the RFID reader allows locating products on a retail shelf with high precision. This facilitates the clustering of products and the distinctive separation of products on different shelves. Having detailed knowledge on what product is where on the shelf supports anticipating out-of-stock situations. The distance-sensitive high frequency RFID system presented in this thesis shows that an increase in RFID location information is technologically feasible. However, this RFID system requires information about a tag’s orientation derived from a tilt sensor for proper operation. Currently, incorporating sensors into passive RFID tags involves significant costs. Further analysis will be required to investigate whether these investments in sensor technology are compensated by the benefits derived from the additional information.
III.3 Weight-sensitive Foam

III.3.1 Introduction

In contrast to RFID, this section presents a class of technology, where the sensors are not attached to items themselves but are embedded in the environment. This approach differs from the previous one in two ways. First, the technology senses the object directly through the object’s physical characteristics, not through an intermediate device (e.g. RFID tag) attached to it. Second, the technology contributes to the system costs only with a fixed part.

There exist several technologies such as computer vision, ultrasound reflection measurements, infrared barriers, and pressure sensitivity that can provide indirect object identification. Research on computer vision has developed powerful algorithms for pattern recognition, but these algorithms still require considerable processing power. Computer vision faces challenges in high dimensional input space to accurately segment and model shapes, in nonlinear mapping, in generalization of shape recognition from a few samples, and in finite memory and computational time. The technology also shows high susceptibility to strong illumination or blocking of view. While illumination is usually well controlled in a retail store, blocking of views cannot be avoided. In order to make maximum use of the shelf space available, products are stocked closely together. Therefore, products are likely to block the view of a significant portion of another product. Additional blocking occurs through the outlines of the shelf and temporarily through customers that may interfere with a camera’s field of vision when standing in front of a shelf or reaching for a product. A large number of cameras could diminish some of these challenges arising from blocking through the combinations of multiple views. However, costs increase significantly with the number of cameras. In addition, image processing in general requires considerable processing power and memory.

Infrared sensors are manufactured at low-cost, but photoelectric barriers, which may be arranged in a matrix around the shelf, only provide limited resolution. This approach provides information on whether a specific row is empty but fails to estimate the remaining stocks. Furthermore, bars that would serve as mount for the photoelectric elements require a robust installation on the shelves which increases integration costs.

Ultrasound transceivers may offer some degree of information about stock quantities due to multiple reflections from objects, but object boundary detection
from ultrasound images remains challenging. Ultrasound transceivers are expensive and come in large form factor making them inapplicable for a large-scale installation on retail shelves.

By contrast, pressure sensitivity shows high potential due to low manufacturing costs, sufficient resolution, low data processing, and simple integration. The two important primitives that are derived from load sensing are weight and position. The weights of objects apply local pressures to a surface [104]. From the intrinsic property of weight either single instances of objects or entire classes of similar objects are identified. The position of an object is the location at which the corresponding pressure applies.

Pressure sensitivity is a well-established technology in a wide range of applications, and sophisticated measurement devices are available [105], [106], [107]. Commercially available systems are designed for specific applications such as biomedical applications and system providers include Novel, XSensors, Measurement Specialties, and TekScan. These systems are optimized for best performance that lowers scalability and increases manufacturing costs, e.g. Novel’s AT-25A pressure-sensitive mat (Figure 26) at the dimensions of 70cm · 30cm · 2cm and 2 sensors/cm² costs approximately $7,500 [108]. Figure 26 also shows Tekscan’s I-Scan, a flexible sensor matrix based on resistive sensors with the dimensions 57cm · 88cm, and Tekscan’s FlexiForce – a thin tactile force sensor [109].

![Figure 26: Novel’s AT-25A [108] and Tekscan’s I-Scan and FlexiForce [109].](image)

The retail industry has different requirements - moderate sensitivity at very low cost for wide-area adoption. In order to keep manufacturing costs low, the design of such sensors must account for large-scale manufacturing processes. Further, the sensor mounting must be flexible for roll-to-roll manufacturing and the sensors need to be mounted at high speed during a continuous casting process.

In this section, weight-sensitive foam is presented. The foam is embedded in the environment and allows monitoring product availability on retail shelves. In the
following subsections, the physical characteristic of such foam is analyzed, a measurements system is presented, and the accuracy of the product detections is evaluated. The section is concluded by cost considerations for the retail industry application and a discussion of the system’s limitations.

### III.3.2 Related work

There are three major classes of sensing principles for thin flexible pressure sensors: piezoelectricity, resistivity, and capacitivity. These measurement principles are very reliable because they are based on direct contact to an object that limits disturbances from environmental influences.

Piezoelectricity describes the generation of a voltage when a mechanical force is applied to crystals and certain ceramics. The piezoelectric effect is reversible and only suitable for the detection of changes in the applied force. This is due to the fact that the leakage current that allows the detection of a change in the applied force, decreases steadily to zero when the load remains constant [105]. The piezoelectric effect does not have a direct current (DC) response that makes it unreceptive to the constant weight of consumer products.

The exposure of semi-conductive material to an applied force leads to a change in electrical resistivity. Weight-sensors based on electrical resistivity are manufactured as thin and flexible sensors with a high spatial resolution. The most widespread technology for piezoresistive sensors uses two thin flexible polymer sheets with screen printed (thick film) or deposited (thin film) conductive lines [105]. The design of resistive sensor arrays are split into two major groups: sensor matrices using electrode contacting on opposite sides of the sensor material and matrices using one-sided contacting [110]. According to Weiss et al., sensor matrices with opposed contacting as proposed by [111], [112], and [113] show high scalability, because the sampling electronics remains simple even for large sensor arrays. However, the measurement suffers from cross-talk between the adjacent stripes that reduces the resolution. In addition, large area sensors limit the sampling speed due to the capacitance of the long electrode stripes that form an RC element together with the resistance of the sensor material. For the design with single-side contacted matrices the capacitive effects do not appear. This allows for high sampling speed. However, this technique requires the connection of each electrode to a multiplexer, which requires connection lines to be routed in between electrodes. This limits the maximum spatial resolution. Additionally, the connection lines lead to large and varying stray capacitances which are difficult to account for.

Although some resistive sensors have single element sensing areas as low as 0.1mm$^2$ and typical scanning rates of 250,000-800,000 sensors per second, these
sensors show high non-linearity and a complex dependence of the response on the number of pressure cycles and history [105]. For some resistive sensors the response is also dependent on temperature and humidity that lead to poor stability and limited durability [105].

The capacitive sensor technology uses a three-layer structure. The conductive rows and columns enclose a pressure-receptive layer that consists of a non-conductive elastomer with high dielectric constant [105]. Capacitive sensors provide higher accuracy than resistive sensors, in which particles are often not uniformly distributed [114]. Capacitive sensors also show better linearity, less affection to the number of pressure cycles and history, and lower dependence on temperature and humidity than resistive sensors. Additionally, capacitive foil sensors may be fabricated with off-the-shelf materials in a simple manufacturing process while resistive sensors require special materials and more complex manufacturing.

Academic research has developed various applications that are based on weight-sensing. Addlesee et al. has developed a weight-sensitive floor based on load sensors to identify and track people [115]. The active floor is a square grid of conventional carpet tiles supported at the corners by cylindrical load cells to provide the total vertical force. The authors use load cells made out of high-strength L168 aluminum alloy, to achieve high robustness of the floor. McElligott et al. has developed “Z-tiles” to be used as control surfaces where force, weight distribution or motion, is used as control parameters [116]. The sensors consist of silicon rubber and carbon granules that show a change in electrical resistance when a force is applied. Schmidt et al. propose the use of load cells in table-legs for context acquisition in an everyday environment [104]. They demonstrate that the position of an object on a surface is accurately and reliably determined. However, the flexibility of their approach is limited to determining the static position of a single new object. In addition, the integration of robust, expensive load cells requires significant modifications to the environment such as tables, furniture, or floors. The applications described above require large, robust, and inflexible sensors.

The “Magic Carpet” by [117] uses piezoelectric wires that offer some degree of flexibility. However, the wire only generates an output when there is a variation in the applied force. Hence, objects that remain still cannot be detected. Research on capacitive sensitivity for flexible mounts has developed smart textiles where conductive fibers are weaved into the fabrics or electrodes are stitched onto textiles [113], [118], [119]. Although these textiles offer high flexibility, this approach comes at significant manufacturing costs. Furthermore, textiles provide instable stands to products with a high center of gravity due to the fabric’s compressibility.
Lumelsky et al. describes the requirements of large-area, flexible array of sensors (sensitive “skin”) with data processing capabilities intended to be used to cover the entire surface of a machine or parts of a human body [119]. Someya et al. report the development of a flexible electronic skin, which is capable of detecting pressure and temperature [120]. In order to sense pressure, they embedded organic transistor-based electronic circuits into a thin plastic film.

Curtis describes a device for detecting intrusion or for inventory control purposes [121]. This device is based on a mat, which has loosely superposed metallic foil sheets separated by a sheet of paper or plastic foil to form a capacitor. The mat shows a change in capacitance when a force is applied. However, Curtis does not describe any clustering of the mat and his system is only capable of detecting the change in capacitance caused by the first object that is placed on that mat. As long as one object remains on the mat, adding or removing other objects cannot be detected. Lazzara describes an apparatus, which detects the removal of an object from a location [122]. This apparatus is based on a compressible detection pad with an upper conducting plate, a lower conducting plate, and an intermediate resilient dielectric layer. The pad shows a compression of the dielectric layer when a force is applied resulting in a change in capacitance.

The retail industry represents a large market with the potential to benefit from enabling technologies that reduce sales loss. However, the industry is hesitant to make larger investments in technology. Therefore, mainly economical factors will decide whether the successful adoption of pervasive computing technology for product availability monitoring continues. The monitoring system presented in this subsection is based on inexpensive polyolefin foam that serves as mount for capacitive sensing elements. In contrast to previous work, this sensor system is designed for roll-to-roll based manufacturing and low production costs.

### III.3.3 The physical characteristics of the sensor mount

A measurements system based on capacitive sensing consists of two parts – the measurement electronics and the sensor. The sensor is formed by two electrodes on opposite sides of a spacer material. The exposure to weight leads to a deflection of the spacer material. This results in a reduction of the gap between two electrodes, and according to Eq. (3) to a deviation in capacitance.

$$\Delta C = \varepsilon \cdot \frac{A}{\Delta d},$$

**Eq. (3)**

where $A$ is the area covered by both electrodes, $\varepsilon$ the dielectric constant of the
polymer, and $\Delta d$ the change in the gap between the two electrodes.

Figure 27: The two electrodes enclosing a dielectric compressible material form a variable capacitor.

The properties of capacitive sensor arrays significantly depend on the physical characteristics of the film that serves as a mount for the capacitors’ electrodes. Soft cellular polymer films are commercially available and widely used for packaging. The low stiffness of polymeric foams makes them ideal for applications in mechanical damping, e.g. seat cushions. In packaging, polymeric foams are also widespread due to their low weight and excellent thermal insulation properties [123].

Polyolefin elastomers are a relatively new class of polymers. Most commercially available polyolefin elastomers are copolymers of either ethylene-butene or ethylene-octene with properties ranging from amorphous to crystalline, and low to very high molecular weight. An inherent characteristic common to all polyolefins is a nonpolar, nonporous, low-energy surface unreceptive to inks, and lacquers without special oxidative pretreatment. The two most important and common polyolefins are polyethylene and polypropylene. These polyolefins are very popular due to their low cost and wide range of applications.

Most polymers can be foamed based on mechanical, physical, and chemical processing [123]. Cellular polymers significantly expand the property range of solid polymers. Polyolefin foams with lens-shape voids show highly anisotropic elastic properties with elastic moduli between 1 and 10MPa. These foams are significantly softer in their cross-sections than bulk polymers with elastic moduli of about 1GPa. The softness of these foams makes them suitable for capacitive sensor networks that require high sensitivity to accurately detect loads [124] [125].

The foam’s thickness and compressibility significantly influences the range of capacitance deviation. While moderate thickness and high compressibility are desirable to achieve high sensitivity for light products, these parameters lead to low resilience and unstable stands for heavy products with a high center of gravity.
From numerous polyolefin foams, polyolefin foam from Sekisui Alveo (TEE0300.25) with a density of 330kg/m³ [126], a thickness of 250μm and a dielectric constant of 1.7 is selected (cf. A.1 for an extensive evaluation). The TEE0300.25 features moderate compressibility and shape recovery for stable stands. Test results for tensile strength (ISO-1926), breaking elongation (ISO-1926), compression load deflection (ISO3386-1), compression set (ISO-1856-C), and shore strength (868-1985, ASTM D2240) suggest good durability [126]. In contrast to non-generic foam that is based on polyvinylchloride or polyurethane, it is stable in its web direction, which is an advantage during conversion (e.g. coating, lamination).

In order to measure the stress-strain behavior of the TEE0300.25, the experimental setup introduced in [127] is used. The experimental setup is schematically depicted in Figure 28. A polytetrafluoroethylene bar with a weight of ~14g is loaded to the sample by means of pressurized air. With this setup, mechanical stresses of up to 1MPa are achieved, which allows the investigation of foams with an elastic modulus Y of up to 100MPa. The foam is surrounded by electrodes as depicted in Figure 28. To determine the sample strain S and the elastic modulus Y, the stress-dependent capacitance C(T) is measured with an LCR meter. Next, the strain S of the sample is determined from the load-dependent capacitance according to a simplified model that consists of a serial connection of polymer layers of thickness s₁ that are separated by a gap of thickness s₂ (Figure 28). The capacitance of the sample is given by:

$$\frac{1}{C(T)} = \frac{s_1}{\varepsilon_0 \varepsilon A} + \frac{s_2(T)}{\varepsilon_0 A},$$

Eq. (4)

where \( \varepsilon_0 \) is the permittivity of vacuum, \( \varepsilon \) the dielectric constant of polymer and A the electrode area under mechanical stress. In Eq. (4), it is assumed that the applied compressive stress T solely results in a reduction of the void size. According to [127], this assumption is adequate as the foam is much softer than bulk PP, and the voids are more easily compressed than the surrounding polymer. Under this assumption, the strain \( S = (s_2(T) - s_2(T = 0))/s \) of the polymer is calculated from the stress-dependent capacitance according to:

$$S = \frac{\varepsilon_0 A}{s} \left\{ \frac{s}{C(T)} - \frac{1}{C(T = 0)} \right\},$$

Eq. (5)
Figure 29 shows the strain versus stress relationship of TEE0300.25 in cross-section. Stress cycles with amplitudes of 0.1, 0.3 and 0.5MPa were used at pressure ramps of 0.3, 0.4 and 0.56MPa/min, respectively [128]. For moderate stress levels of up to 0.1MPa, the stress-strain relationship is nearly linear. However, nonlinearities and hysteresis are distinctive when higher stress levels are applied as demonstrated by the cyclic deformation measurements with amplitudes of 0.3 and 0.5MPa, respectively. When pressure of up to 0.04MPa is applied to the foam, an elastic modulus of approximately 4.4MPa is derived from the slope of the stress-strain curve. Figure 30 shows the small signal elastic modulus of the foam when pressure is steadily increased to 0.5MPa. The small signal elastic modulus is derived through numerical differentiation of the smoothed signals illustrated in Figure 29. On a logarithmic pressure scale, three distinctive regions are identified that characterize the elastic properties of foams. At low pressure levels, foam behaves like linear elastic solid (I), while the elastic modulus is highly nonlinear at higher stress levels indicating a softening region (II) and a densification (III) region. For pressure measurement applications, foam is only used in the linear region because strong nonlinearities and hysteresis are observed in the softening and densification region that make it difficult to conclude applied pressures. In the linear region, the elastic modulus of the foam is more than two orders of magnitude smaller than the bulk modulus of polymers. This result is rather unexpected because a large number of closed-cell foams show linear scaling of the relative elastic modulus versus relative density [129]. The cross-section image of an electron microscope scan (Figure 31) shows that the foam consists of only few large and anisotropically-shaped voids across the foam’s thickness. These voids account for the high elasticity in the foam’s cross section because the elastic modulus of the foam largely depends...
on the bending characteristics of the two-dimensional cell walls. Therefore, the scaling laws derived for small and isotropically-distributed spherical voids do not apply [128].

Elasticity, hysteresis, and reversibility are important limitations of all flexible thin pressure sensors. The physical characteristics of this foam show suitable elasticity at high linearity and low dependence on the number of pressure cycles. Theses characteristics make it suitable to serve as spacer material for the capacitive sensor network. The foam’s weight sensitivity is determined by loading a 1cm² capacitive sensor element with different weights. The capacitance between two electrodes is measured with an Agilent 4285 LCR meter. As shown in Figure 32, a linear dependence is obtained between the change in capacitance and the load.

![Figure 29: Illustration of strain vs. stress of polyolefin packaging foam in cross-section. The graph reveals a linear relationship between strain and stress at low stress levels of up to 0.5 bar; nonlinearities and hysteresis occur only at higher stress levels.](image-url)
Figure 30: Illustration of small signal elastic modulus of the foam vs. pressure when pressure is increased up to 0.5 MPa at 0.56 MPa/min. Three distinct regions characterize the elastic behavior of the foam: linear (I) and nonlinear elasticity regions with softening (II) and densification (III) of the foam.

Figure 31: The cross-section electron microscope scan of polyolefin packaging foam shows few large and anisotropically shaped voids across the thickness of the foam.
III.3.4 Weight-sensitive sensor

The passive array of capacitors is composed of electrode leads that run in parallel across the foam on both sides but in perpendicular directions. This design is important to allow for roll-to-roll based manufacturing that results in low prices per square meter.

The cross-over points of the leads form the capacitive elements (Figure 33) [113], [130]. The weights of the products lead to local deflections of the foam. Such a compression reduces the gap between electrodes on opposite sides and results in an increase in capacitance. Accordingly, the number of products on a shelf is determined through the detection of diverging capacitances.

Figure 32: The change in capacitance of a single sensor element correlates with the load.

Figure 33: The leads run in parallel across the foam and form capacitors at their crossover points.

The graph on the upper left of Figure 34 shows the deviation in capacitance when
a weight of 11g is placed on a 10mm \times 10mm electrode mounted on 250\mu m-thick TEE0300. The weight leads to a steep edge representing a change of about 25fF. When the weight is removed, the capacitance immediately returns to a value that is 2.5fF higher than its original value. On the lower left of Figure 34, the experiment is repeated with 200\mu m-thick TEE0400. It shows a slower rising edge that reaches a change of 15fF. When the weight is removed, the capacitance drops quickly but returns to a value 4fF higher than the original value. It is assumed that the weight leads to a permanent deformation of the foam. The same effect was observed to be more severe with softer and thicker foams. On the right, measurements are taken of the capacitance of 5mm \times 5mm electrodes that are mounted on 200\mu m-thick TEE0300 foam. The weight is steadily increased and decreased from 0g/cm^2 to 100g/cm^2 resulting in a deviation in capacitance of 50fF. The measurement hardware is capable of resolving up to 4fF that corresponds to a weight resolution of 8g/cm^2. However, the noise level limits the actual resolution to about 5fF, which corresponds to 10g/cm^2. Additionally, the graph for 200\mu m-thick TEE300 shows high resilience (variations remain below the noise level) and sufficient sensitivity. In addition, crosstalk and creep do not exceed the noise level for this design where two neighboring electrodes are separated by 20mm. Therefore, this combination of 200\mu m-thick TEE300 foam with a sensor grid of 20mm and electrode pads of 25mm^2 meets the requirements of the retail application.

Figure 34: Shows the deviation in capacitance for TEE300.25 and TEE400.2 with a load of 11g on 10mm \times 10mm electrode pads (left); and the deviation in capacitance according to a linear change in weight from 0-100g/cm^2 for TEE0300.2 on a 5mm \times 5mm electrode pad (right).
This design, where the leads not only form the capacitors but also connect several electrodes, reduces the required amount of connection lines (Figure 35) [113]. Therefore, the maximum resolution, that corresponds to the minimal gap between two parallel input leads, is not limited by connection lines that require routing between electrodes. In this first design, the leads are hand-painted with conductive silver that show a resistance of approx. 2Ohm/meter. However, during manufacturing, a simple continuous casting process would superimpose the leads on the foam.

Figure 35 shows foam with a size of 30cm · 20cm containing 96 capacitive sensors in 12 columns and 8 rows. Each electrode is 5mm wide and separated by 20mm from the next electrode.

![Figure 35: Conductive silver leads run in parallel across the flexible polyolefin foam to form the 96 sensors. When multiplexed, the leads on different sides of the foam form the capacitive elements.](image)

III.3.5 Measurement system and data processing

The capacitance is a measure of the amount of charge Q stored on each plate for a given potential difference or voltage V that appears between plates. Measurement methods for capacitances are arranged in four main categories: resonance, oscillation, bridge and charge/discharge [131]. The resonance method is able to measure both the unknown capacitance and its parallel loss over a wide frequency range [132]. However, this method is not suitable for continuous measurements because the operating steps are carried out manually. The accuracy of the RC oscillation methods is insufficient – the oscillation frequency shows susceptibility to the shutting conductance, poor sensitivity to small capacitance changes and poor frequency stability [132]. Although LC oscillation methods work with frequencies of up to 200MHz, there is a baseline drift due to the stray capacitance.
Both the resonance methods and the oscillation methods have the disadvantage of having the stray capacitance being included in the measurements. Although techniques exist for charge/discharge methods that are robust to stray capacitance, the AC bridge with feedback is recommended for measurements using frequencies below 100KHz [132].

In this design, a sigma-delta capacitance-to-digital converter is used because it is available in small integrated circuit (IC) packaging. The standard sigma-delta analog-to-digital converter (ADC) is implemented by switching on-chip fixed capacitors and balancing the charge between a variable voltage input and a defined voltage-reference input. The output data represents the ratio between the sensor capacitance and the reference capacitance.

The electronics contain an AD7745 Sigma-Delta converter from Analog Devices that measures the capacitances. This converter offers high accuracy of 4fF with an update rate of 90Hz by direct conversion of the capacitance to a digital signal. A multiplexer circuit switches sequentially row by row through all the leads. While the capacitance between two leads is measured, all other leads are grounded to minimize crosstalk and to keep the stray capacitances at a stable value. Parasitic capacitances are compensated directly in the converter.

The system is initialized by measuring each sensor in an unloaded state to compensate for variations in capacitances due to inaccuracies in manufacturing and to different lengths of the connection lines. These individual values are stored to serve as reference values. Further measurements are compared to the individual reference values, and if a measured sensor exceeds its reference value by a certain threshold value, the system assumes the presence of an object. Once an entire measurement cycle is completed, all loaded and unloaded sensors are displayed. For a prototype application, the color of loaded pads changes. To determine the quantity of stocks on a shelf, the loaded sensors per row are summed up. Products are not necessarily displayed in a consecutive order but may be arbitrarily distributed. If the number of loaded sensors per row is small, the shelf needs to be replenished. The application uses a graphical interface to inform the staff of low stocks (Figure 36). Other ways to alert the personnel such as turning on a light above the shelf or sending a text message to the cellular phone of the sales floor manager may also be applicable.
III.3.6 Energy consumption

The number of sensor measurements per time period significantly influences the power consumption of the system. While a low update rate preserves energy but results in a possible long delay until an out-of-stock situation is detected, a high update rate may require frequent replacing of the batteries that power the system. Taking into account that the actual replenishment process requires a considerable amount of time – the staff has to respond to the system’s alert, move the merchandise from the backroom to the dedicated location and stock the shelf – the update rate of the monitoring system will be chosen between 5min and 30min. Table 2 lists the energy consumption of the individual electronics components for two modes of operation, busy and idle, and gives the number of components used in this design.
Table 2: Energy consumption of the electronic components (measurements).

<table>
<thead>
<tr>
<th>Component</th>
<th>Busy [mW]</th>
<th>Idle [mW]</th>
<th># of Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSP430</td>
<td>0.5</td>
<td>0.06</td>
<td>1</td>
</tr>
<tr>
<td>AD7745</td>
<td>2.3</td>
<td>0.002</td>
<td>1</td>
</tr>
<tr>
<td>ADG709</td>
<td>0.005</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>HEF4514B</td>
<td>0.4</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>MM74HC</td>
<td>0.05</td>
<td>-</td>
<td>24</td>
</tr>
<tr>
<td>ZigBee (1.8V)</td>
<td>34.2</td>
<td>0.01</td>
<td>1</td>
</tr>
</tbody>
</table>

The AD7745 operates at 80Hz. Therefore, 2048 sensors are measured in approximately 30s. In order to compute the power consumption of the system for cycle times of 5min, 10min, 15min, and 30min, it is assumed that the system is busy for 30s to perform the measurements and remains in the power-saving idle mode for the remaining time. The system’s power consumption for the different cycle times is given in Table 3.

Table 3: Power consumption for different cycle times.

<table>
<thead>
<tr>
<th>Cycle time</th>
<th>Power consumption [mW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>5min</td>
<td>0.69</td>
</tr>
<tr>
<td>10min</td>
<td>0.3809</td>
</tr>
<tr>
<td>15min</td>
<td>0.2746</td>
</tr>
<tr>
<td>30min</td>
<td>0.1683</td>
</tr>
</tbody>
</table>

Two AAA batteries at 3V have a capacity of about 2310mAh, and one of 2 AA batteries is about 5000mAh. Therefore, a system with a cycle time of 30min is powered for 10.3 years. However, these calculations do not account for the energy consumption of the wireless module that transmits the notification about an out-of-stock situation along with the sensor data to a backend system. The system only uses the wireless link if an out-of-stock situation occurs but the number of occurrences is difficult to estimate. A lifetime of 10.3 years at a cycle time of 30min represents an upper bound because it assumes that the wireless link is never used. Under the assumption that an out-of-stock situation is detected each time the system updates the measurements, the total power consumption of the system including a wireless data transmission is 0.748mW. This results in a lower bound for battery lifetime of 2.3 years.

Out-of-stock rates for products are given in the range of 5-10%. If an out-of-stock rate of 7% is assumed for an average American retailer that carries 45,000 SKUs,
3,150 SKUs are out-of-stock at a time. The 45,000 SKUs are displayed on approximately 15km of shelves. The weight-sensitive foam covers up to 2m in shelf length. Therefore, 7,500 individual weight-sensitive foams are required to cover the entire shelf space. It is concluded that weight-sensitive foam detects an OOS situation in average every 2.38 days. Consequently, the wireless channel is only used every 114 cycles. This results in battery lifetime for 2 AAA of 4.6 years and for 2 AA of 9.96 years.

### III.3.7 System evaluation

Most retail stores offer an enormous variety of products. In order to test the design with a limited amount of products, the evaluation is restricted to products that show significant stock-out rates. According to [35], hair care products show highest OOS rates with an average of 9.8%. These products also account for the largest share of estimated retail loss of 4.5%. Out-of-stock rates for hair care products are followed by laundry products (7.7%), diapers (7.0%), feminine hygiene products (6.8%), toilet tissues (6.6%), and salty snacks (5.3%). From these categories, products are chosen to represent a specific category and to simulate a challenge for detection. Therefore, the design is tested with products that show significant weights on small footprints and light weights on large footprints. The characteristics of these products are listed in Table 4.

**Table 4: Shows the weight of the products, the covered area on the foam, the footprint as seen by the sensors (actual pressure area), and the weight per area.**

<table>
<thead>
<tr>
<th>Product</th>
<th>Absolute Weight [g]</th>
<th>Covered Area [cm²]</th>
<th>Footprint [cm²]</th>
<th>Weight [g/cm²]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shower gel</td>
<td>260</td>
<td>2.5 x 6.5</td>
<td>4.8</td>
<td>54</td>
</tr>
<tr>
<td>Laundry</td>
<td>2200</td>
<td>8 x 17.5</td>
<td>54</td>
<td>41</td>
</tr>
<tr>
<td>Diapers</td>
<td>3000</td>
<td>10 x 22</td>
<td>140</td>
<td>21</td>
</tr>
<tr>
<td>Tissues</td>
<td>220</td>
<td>7 x 11</td>
<td>28</td>
<td>8</td>
</tr>
<tr>
<td>Peanuts</td>
<td>240</td>
<td>4.5 x 6.5</td>
<td>5.5</td>
<td>44</td>
</tr>
</tbody>
</table>

The product availability monitoring system does not aim to determine the explicit number of products on the display area. Instead, the personnel are informed only when a small number of sensors per row are loaded because this corresponds to an almost empty shelf (Figure 36). Therefore, it is critical that the system accurately detects and displays the sensors that are exposed to the weights of products. Errors occur when completely or partly covered sensors do not report a load (false negatives) or when sensors show the detection of a product when they are not
loaded (false positives). The error rate for false negatives and false positives is determined by first arranging the previously selected products on the display area and then removing individual items in three different ways – front-to-back, back-to-front, and randomly. After each removal, it is visually determined which sensors are covered and the results are compared to the actual output of the system. Ten test runs are performed for each product and method of removing items (results are listed in Table 5). The tests for shower gel show that 1.1% of all covered sensors do not show a load (false negatives) and 0% of the sensors report a weight detection even though they are not exposed to any weight of a product (false positives). For laundry bags, the same test shows 2.8% false negatives and 0% false positives, and for diapers, the results are 7.0% false negatives and 1.2% false positives. Peanuts (salty snacks) show the highest error rates with 9.7% of false negatives and 4.1% of false positives. The declared weight per area for tissues is around the minimal resolution of the system, which leads to inaccurate detection. However, in retail stores tissue boxes are usually stockpiled. A minimum of two stacked boxes are reliably detected by the hardware with an error rate of 2.5% of false negatives and 3.3% of false positives. Note that customers usually take products from the front so that the products in the back of the shelf remain piled. The sensors’ resilience is tested by first loading each sensor and then measuring the false negatives and false positives after the complete removal of all loads. The design shows high robustness with an error rate of 0% for both. This evaluation of the detection of selected product shows overall error ranges of 1.1%-9.7% for false negatives and 0%-4.1% for false positives.

Table 5: The table shows the false negatives and false positives for the tested products.

<table>
<thead>
<tr>
<th>Product</th>
<th>False Negatives</th>
<th>False Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shower gel</td>
<td>1.1%</td>
<td>0%</td>
</tr>
<tr>
<td>Laundry</td>
<td>2.8%</td>
<td>0%</td>
</tr>
<tr>
<td>Diapers</td>
<td>7.0%</td>
<td>1.2%</td>
</tr>
<tr>
<td>Tissues</td>
<td>9.7%</td>
<td>4.1%</td>
</tr>
<tr>
<td>Peanuts</td>
<td>2.5%</td>
<td>3.3%</td>
</tr>
</tbody>
</table>

False negatives result in early triggering of the replenishment process. False negatives are mainly the result of uneven footprints where the covered area is not equal to the area that pressure is applied to. In particular, peanuts, which come in small bags, show uneven footprints. This explains the significant number of false negatives that occurred. All other products show moderate error rates of false
negatives. However, the more significant measure for the evaluation of the system is the error rate of false positives. False positives lead to a higher recorded inventory level than actual inventory level. Therefore, the system risks not anticipating out-of-shelf situations because it records a stock quantity that is too high. Nevertheless, preliminary tests show moderate error rates for false positives. In addition, these inaccuracies can be accounted for by increasing the threshold value at which the system concludes an almost empty shelf. Therefore, the results of this evaluation show that the system can accurately anticipate out-of-shelf situations and inform the personnel before an OOS situation occurs.

The reduction of out-of-shelf situations in retail stores is a multi-tier problem. The system only provides visibility to the quantity of stocks on retail shelves, but cannot account for slow replenishment by the personnel. However, in order to compensate for slow response time the threshold value at which the system displays the replenishment notification can be increased. Since the threshold value is hardware independent, it is adjusted after installation at the retail store. To determine the optimal threshold value and to exactly determine the reduction of out-of-shelf situation, tests at retail stores are required where shelves are completely equipped with weight-sensitive foam. The in-store shelf testing is also necessary to prove the durability of the system.

The design offers potential to increase the resolution by narrowing the gaps between leads. A higher number of weight sensors on a smaller area not only determines the exact weight of a product, but also permits deriving its footprint. Therefore, current research intends to incorporate pressure pattern analysis to determine individual footprints. This will allow deriving the exact number of products on a shelf.

Misplacements account for sales losses of approximately 2% [63]. A system that detects weight and footprint of individual products can locate misplaced items if their weights and footprints differ from those of the other products on the shelf. However, the detection granularity is often limited to product categories because many products are sold in standardized packaging (e.g. cans). Thus, a system based on weight and footprint sensing will not significantly reduce the number of misplacements.

**III.3.8 Cost considerations of manufacturing processes**

Economic considerations of system costs are one of the most critical factors for successful adoption of monitoring technologies in the retail industry. In order to estimate the costs of the monitoring system, the individual costs are calculated. These costs include material cost for polyolefin foam, conductive silver, and
protective layers, set-up costs for the manufacturing machines, and costs for the electronics and power supply (note, integration costs significantly depend on the backend inventory system and are not considered here). Manufacturers give prices for polyolefin foam of about $20/kg. This results in $0.1/running meter for foam with a density of 25kg/m$^3$ and the dimensions of 1m $\times$ 0.6m $\times$ 200e-6m. Conductive silver comes in containers of 60ml (90g) at $90. If leads are printed according to the grid suggested above, 40 leads on each side of the foam with widths of 5mm and lengths of 0.6m (depth of a retail shelf) are required. Manufacturers suggest that current casting processes achieve lead thicknesses as thin as 4μm-20μm for carrier materials with a thickness of 200μm. This results in a volume of 9.6e-7m$^3$ to 4.8e-7m$^3$ of conductive silver per running meter or costs of $1.4-$7.2. Prices for protective layers of PE foils are about $0.1/running meter. The manufacturing of the sensors requires approximately 4 different processes. Each process comes at set-up costs of about $10. However, it is assumed that at least 100m of foam are processed at once in roll-to-roll based manufacturing. Therefore, the total set-up costs are concluded to be $0.4 per running meter. Chip manufacturers quote prices for chips that include measurement electronics and WLAN modules of $5 [133]. It is important to note that a single chip covers an area of at least two square meters. Thus, the electronics contribute $2.5 to the total costs. Two AAA batteries, priced about $1.5, are expected to be sufficient to power the system for years. The total of these individual costs results in system costs of $6.1-$12.1 per running meter. It is expected that prices will drop due to large-scale production.

The system costs are further reduced if copper with a thickness of 100nm is used instead of conductive silver. With this modification, the costs for the sensors will drop down to $0.9 [134], and the total system cost will result in $5.3. However, in this approach the leads cannot be printed on the foam because its surface roughness exceeds the thickness of the leads. Instead, the leads will have to be imposed on the PE foils that cover the foam for protection. The influence of the lower conductance of copper on the system has not been evaluated in detail but preliminary tests indicate that the effect on the system is negligible.

Because the costs of the proposed systems are significantly lower than those for manual data collection and RFID, the design may meet the economical requirements of the retail industry.

1 $6.1-$12 (total costs) = $1.44-$7.2 (conductive silver) + $0.1 (foam) + $0.2 (PE foil) + $0.4 (set-up) + $1.5 (batteries) + $2.5 (electronics)
III.4 A Qualitative Comparison

Automatic product monitoring intends to replace the time and cost intensive process of manually collecting information about shelf stocks. The automation of this process allows for high update rates on product availability, creates data that supports the store manager in keeping the shelves stocked, and offers high product availability to consumers. As discussed above, the technologies reviewed use fundamentally different approaches: RFID uses an object-centered approach and weight-sensitive foam uses an environment-embedded approach. This subchapter discusses the advantages and disadvantages of both approaches based on the data they generate, the barriers to overcome for fast adoption, additional benefits derived from deployment, and possible consumer concerns related to the technologies.

III.4.1 Information granularity

The degree of RFID deployment significantly influences the level of detail at which inventory information is available. As discussed above, RFID on case and pallet levels helps separating backroom inventory from sales floor inventory. However, RFID on these levels fails to provide detailed information about stock levels on individual shelves. Taking into consideration that cases and pallets will already be equipped with RFID tags to optimize logistics processes along the supply chain, making use of the same tags to improve the in-store inventory system is relatively straight-forward. Infrastructure investments are required for gate readers at the door that separates the backroom from the sales floor. In addition, investments are necessary for readers to capture RFID information at the delivery dock and the waste disposal station. Additionally, the data has to be integrated into existing inventory system to be made available.

The introduction of RFID at case and pallet levels aims to improve the efficiency of creating a record of deliveries made to a store and reducing the amount of incorrect deliveries. It also increases the data accuracy at the stock unit level and reduces the absolute difference between the system record and the physically available inventory. In turn, the service level improves due to fewer out-of-stock situations.

Despite the improvements in data accuracy through the deployment of RFID at case and pallet level, the exact product quantities remains concealed and on-hand inventory remains subject to estimations. Shrinkage and misplacements may still deteriorate inventory records, which can only be updated through physical audits.

Item-level tagging, as opposed to tagging of cases or pallets, is a means of bringing the sensors directly to the shelves. Reader antennae will be embedded in non-
conductive shelves to directly detect the items on display. Each item will be explicitly identified and matched against the system records. In order to distinguish between different shelves or even different parts of a shelf, distance-sensing may be incorporated. However, distance-sensing requires the integration of a tilt sensor into the RFID tags that significantly increases costs.

Inventory management may be further improved by combining RFID with sensors at item level, e.g. temperature sensors, humidity sensors, vibration sensors, etc. Detailed information about the shipping environment is of special interest for perishable goods in global supply chains. Control and visibility over product handling is limited due to several echelons of the supply chain. Currently, battery-powered devices are used to monitor the shipping environment of goods in a cold chain but the cost of these devices, their bulkiness, and their limited lifetime prevent high market penetration. As a result, only limited information is available about the cold chain, which precludes useful insights as to its efficiency.

The availability of cheaper monitoring devices will result in increased market penetration leading to more detailed information about the shipping environment. This will support the analysis and improvement of a supply chain’s efficiency resulting in reduced shrinkage and extended shelf life [135].

The deployment of an environment-embedded approach such as weight-sensitive foam does not compete with current deployments of RFID at case and pallet levels. On the contrary, it is viewed as the consistent continuation of increasing product visibility down to the shelf level. It is a replacement of the more expensive RFID item-level tagging, but still provides sufficient information about shelf stocks.

Weight-sensitive foam is a pervasive product availability monitoring system that anticipates out-of-shelf situations and informs the personnel when a particular shelf needs to be replenished. This monitoring system is based on force-receptive foam that allows inexpensive roll-to-roll based manufacturing. Preliminary tests indicate that the system offers sufficient sensitivity to reliably detect low stocks of products. The error rates of false negatives and false positives are moderate and can be compensated by increasing the threshold value at which the personnel is informed.

However, this system is neither capable of determining the exact number of products on display nor identifying the product types. It monitors the loaded shelf space in comparison to the total allocated shelf space and anticipates out-of-stock situations if the loaded area becomes very small. Hence, the store manager will not receive a request to replenish a shelf with a certain amount of items, but is informed about depletion. Based on this information, the store manager will direct the staff to
that particular shelf. The shelf will be replenished from cases and boxes without a priori knowledge of the exact amount required.

Anticipating out-of-stock situations based on loaded shelf space makes this approach highly scalable. The system does not require any information about the products (e.g. weight, footprint) that are on display on any particular shelf, and therefore, forms a stand-alone system. However, this comes at the cost of not having explicit knowledge about the quantity and type of the monitored product. Information about product quantities on shelves can be detailed by increasing the sensor density. Higher resolution of the pressure patterns leads to more accurate detection of footprints. This allows distinguishing, and to some extent, categorizing products. In contrast to the relative detection of product availability through loaded shelf space, identifying individual products provides information about the explicit number of remaining products on a shelf. The loaded and unloaded sensors are the building blocks for a pressure pattern similar to pixels which form a digital picture. Computer vision has developed powerful algorithms to detect, cluster, and identify sets of pixels. By using similar algorithms to process the sensor data derived from the weight-sensitive foam, individual loaded sensors are allocated to patterns. These patterns are analyzed to determine the product [136].

In contrast to RFID, weight-sensitive foam provides information about the location of products on a display area. From a marketing perspective, it is of interest to have information about how products are displayed on a retail shelf. While products are usually neatly arranged for a flawless facing when shelves are replenished, customers will disrupt the facing by taking products from a shelf. In order to continuously provide an attractive facing, it is critical to derive information about disarranged products so that personnel can be instructed to fix specific facings.

**III.4.2 Technology adoption**

Management research on the diffusion of new technologies has identified various factors that have a significant influence on the adoption process in a company. Factors influencing IT adoption can be categorized into innovation (technology) and organizational characteristics [137]. Detailed evaluation of both categories identifies standards (e.g. to support compatibility), costs of adoption, technological performance, expected benefits and risks, complexity of adoption, top management support, and coercive influences as the most significant factors [138], [139].

Compatibility and interoperability are critical for all systems deployed in a network of independent partners. The exchange of data among supply chain partners is the major benefit derived from RFID. The exchange of data requires standardized data formats and interfaces. The monitoring of shelf stocks with weight-sensitive foam
is limited to the sales floor. Therefore, the adoption of weight-sensitive foam represents a single echelon solution where this technology is deployed as a proprietary system. Such a single echelon solution suggests showing a significantly faster adoption than a multi-echelon one because it is independent from the cooperation with suppliers and logistics providers. Additionally, weight-sensitive foam allows for partial installation. For example, promotional shelves that are more likely to show out-of-stock situations are usually more noticeable to the consumers.

The performance of the technology is critical. Data capturing must be reliable and the data accurate. Low read rates, interferences, and shielding represent the Achilles heel of RFID. In addition, inaccurate detection of loaded sensors deteriorates the performance of the weight-sensitive foam.

Due to the fact that weight-sensitive foam is a proprietary system with moderate requirements for data processing, the complexity of its installation and operation is sufficiently low. On the contrary, RFID requires experienced personnel for the installation of the infrastructure and connection to the backend system. The increased complexity suggests a lower adoption rate. In addition, RFID is more invasive to the current replenishment process than weight-sensitive foam.

The lack of expertise, the complexity of the technology, and the uncertainty of the technology are risk factors that constrain adoption [140]. Clear implementation guidelines and strategies could lower these risks.

The expected benefits of introducing a new technology must outweigh the deployment and operational costs as well as the risks associated with the investments. Depending on the level of adoption, RFID shows moderate to significant operational costs for transponders. Additionally, significant risks and costs are associated with the replacement of a functional and well-established technology such as the barcode. Weight-sensitive foam suggests lower expected benefits (limited to shelf stock monitoring), but also has lower installation and maintenance costs. The risks associated with this investment are lower because a foam-based solution will not replace a current technology, but may improve the current replenishment process. Furthermore, a solution based on weight-sensitive foam allows for a gradual adoption where foam is first installed on promotional shelves (shelves with highest OOS rates) or for certain product categories before the entire store is equipped.

III.4.3 Further benefits

The benefits arising from a RFID deployment, that aims to improve the replenishment process and product availability, represent only a small share of the overall benefits possibly achieved with RFID. In most supply chains a product
passes through several echelons until it reaches its final destination. RFID on pallet and case level significantly increases the visibility of the flow of goods through such a supply chain and allows for better management and optimization, improved inventory accuracy, reduced costs for logistical operations, higher inventory turnover, and lower safety stocks.

The deployment of RFID at item-level offers further benefits such as lower transaction errors, shrinkage reduction, inventory reduction, labor cost savings, precise location of merchandise, detection of misplacements and unsaleables, more targeted marketing, and speed-up of checkout [141], [142].

Shrinkage may account for up to 1.75% of sales [85]. Seventy-seven percent of retail shrinkage is caused by theft [82]. Tellkamp et al. reports on a Swiss retailer who manually checked for inventory accuracy for razor blades and condoms over a period of three months [143]. The company found that out of 100 products delivered to the store, only 32 razor blades and 20 condoms were scanned – the rest somehow disappeared. RFID item-level tagging potentially reduces theft significantly. RFID at item level may reduce shrinkage as well. IBM estimates the reduction of 47% for retailers and 66% for manufacturers [86] (cf. III.2.2 for more details).

Fully automated checkout systems accelerate the payment process. These systems represent a major benefit since congested check-outs are reported as one of the most negative aspects of supermarket shopping [144], [145]. In-store customer marketing (ICM) utilizes RFID item-level tagging to generate in-store consumer product profiles with the subsequent in-store promotion of secondary product accessories [92]. Sellitto et al. report that companies find benefits from RFID implementation across various business functions that are associated with in-store customer marketing [92].

Furthermore, retailers expect reduce labor costs on activities that currently require significant manual intervention and that will become semi or fully automated. In addition to shelf replenishment, activities such as the receipt of goods, locating merchandise, inventory audits, picking list generation etc. could significantly benefit from RFID. AT Kearney estimates labor cost savings of 9% for manufacturers, and 7.5% at retail stores and warehouses [83], [146]. Accenture estimates the savings for the receipt of goods of 6.5%, while labor in physical inventory count will be completely eliminated [147]. Estimates made by McKinsey show expected savings of 0.5-1.6% in distribution, and 0.9-3.4% at the stores [88]; and Accenture report labor savings in receipt of 5-40%, for stocking 22-30%, 95% for audits, and 5-45% at the check-out [87], [148]. METRO Group estimates the reduction in labor at retail stores at about 17% [84].
Even though the introduction of RFID will increase store occupancy costs, retailers hope that the overall benefits will overcompensate these costs. The absolute benefits arising from item-level tagging are difficult to estimate and vary significantly among studies. Lee et al. suggests that the estimates on labor savings are solid because they are based on detailed time-motion studies that allow deriving accurate labor cost savings, while other estimates may show larger variances [63]. Retailers have yet to implement this technology on a large scale to run extensive pilots and trials that will allow for more accurate forecasts. In contrast to RFID at item level, which offers potential to create added value in many other areas at the retail store and throughout the supply chain, the benefits derived from weight-sensitive foam is limited to the possible reduction of out-of-stock situations and safety stocks as well as labor cost savings in inventory monitoring.

III.4.4 Consumer privacy concerns

III.4.4.1 Introduction
RFID at item level is designed to pervasively track individual objects, which legitimates consumer concerns regarding threats to privacy. The ability to permanently and unrecognizably collect RFID data forms a threat to many individuals of loosing control over personal information [149]. The common fear is that tagged objects may allow unnoticeably establishing a connection to personal data beyond the point of sales. Disclosure of purchase behaviors and misuse of data are among the biggest fears [80]. One specific threat is profiling, which is the collection of information linked to personal data and its classification into expected behaviors [150]. Other threats are the identification of individuals who could be targeted for marketing purposes or the possibility of global surveillance in cases of a large RFID network where people could be tracked as they go about their days.

When Mark Weiser envisioned that computing devices will be embedded everywhere in the environment in a way that they can operate unobtrusively, he also acknowledged that the invisible nature of those devices will make it difficult to know the entities that are in control, the network connections among devices, and the locations where information is collected [151]. The antagonism between the requirements for control and privacy versus usability and performance is well illustrated by the privacy concerns associated with the deployment of RFID technology in supermarkets and retail outlets [152].

In [153], Floerkemeier et al. present an approach that addresses these privacy concerns by integrating a subset of the widely accepted fair information principles
[154] into the communication protocols between RFID readers and tags. The authors contend that having RFID readers to explicitly declare the scope and purpose of their tag data collection, as well as disclosing the identity of their operators, will allow both consumers and regulators to better assess and control the effects of everyday RFID encounters. To display and log the meta-information that is broadcasted over the RFID communication channel, the concept of a watchdog tag is presented that allows privacy-concerned individuals to judge whether a particular RFID reader deployment complies with the corresponding regulations.

In that paper, the concept of a watchdog tag is only demonstrated on a PDA that receives the meta-information about the on-going data collection via a wireless LAN connection. The main contribution of the approach presented in this thesis is a battery-powered watchdog tag that is capable of decoding and displaying the meta-information transmitted by the reader over the RFID communication channel. This prototype represents the first step towards a user study. The study will evaluate the concepts proposed in [155] and compares the suitability of the watchdog tag to other privacy enhancing techniques such as disabling the tag or blocking the tag-to-reader communication.

III.4.4.2 Related work

There is a large variety of privacy enhancing technologies available. These technologies include concepts that rely on disabling the tags permanently [156], [157] or at least temporarily [158], but also approaches that block the tag-to-reader communication [159]. For the watchdog design, the focus is exclusively on the technique proposed by Floerkemeier et al. [153].

There have been other efforts to decode RFID reader commands and to reply to the reader with data. In [160], Ignatov describes an elegant way to build an LF tag with a PIC microcontroller that requires very few discrete components. Carluccio et al. developed a RFID reader command detector that works based on the ISO 14443 protocol, which is primarily used for contact-less smartcards [155]. However, their implementation uses a ready-made analog front-end transceiver chip that demodulates and decodes the reader data. Unfortunately, there is no chip available for the long range HF protocols, such as ICODE1 and ISO 15693. At UHF, Smith et al. presented the WISP platform that emulates an EPCglobal Generation1 Class 1 tag [161], but features additional sensors, such as a light sensor. The WISP platform operates without an additional battery to power the microcontroller.

III.4.4.3 Integrating the fair information practices into RFID protocols

Passive high frequency tags, that operate at 13.56MHz, are widely used to equip products in the consumer goods industry because of their small form factor, low
cost, and resistance against interference. With the high adoption rate of RFID tags in the consumer goods industry, customers are more and more likely to carry around items that are equipped with RFID tags. These tags continue to respond to inventory scans of RFID readers after purchase; mostly without the awareness of the consumer.

Current RFID readers neither offer identification nor provide information about the purpose of the interrogation. Although such an anonymous scan allows for high performance by keeping the exchanged data to a minimum, it does not satisfy the principles of openness and accountability. To meet the requirements by the fair information policy, Floerkemeier et al. suggest including a unique reader policy ID along with a purpose declaration and collection type into the inventory command of a reader (Figure 37).

Figure 37: Shows the existing inventory command (above) and the proposed command extensions (below).

These recommendations are incorporated into the proposed watchdog design. In order to test the design, however, there are no modifications made to the existing reader protocol to avoid firmware updates. Instead, the reader policy ID, the purpose declaration, and the collection type are included in the data fields of the write command. The watchdog tag does not execute the write command but extracts the meta-information and reports it on a display for the user’s interest (Figure 38). However, current passive tags do not provide interfaces for additional devices, and therefore, a passive RFID tag is built with discrete components that allow attaching a display.
**III.4.4.4 Watchdog tag**

The watchdog tag is based on the platform developed in subchapter III.2.4.4. In addition to that system, the watchdog tag contains a display to make the meta-information accessible to the user. The dot-matrix display contains a HD44780 controller. The controller allows addressing 2 lines, on which 20 characters each can be displayed (Figure 38). The LCD display consumes about 2mW. It is unrealistic to build a passive watchdog tag that harvests all its energy from the magnetic field of the reader and operates at ordinary distances rather than at very close proximity to the RFID reader. Therefore, this design incorporates a battery.

However, for daily use a watchdog tag should be incorporated into a cellular phone that already contains a battery and a high resolution display. In addition, this incorporation would allow translating reader identifiers into more meaningful information about the operator using the long-range communication capabilities as mentioned in [155].

![Image of the display of the watchdog tag showing the decoded reader policy ID.](image)

**Figure 38: The display of the watchdog tag shows the decoded reader policy ID.**

**III.4.4.5 Discussion**

In vicinity of the reader, the watchdog tag successfully decodes the meta-information included in the modified write command of an ICODE1-compliant RFID reader. The tag reports the reader policy ID, the purpose declaration, and the collection type on a display for the inspection by the user. It is desirable that a watchdog tag achieves a longer read range than ordinary passive tags that may be attached to items a user carries along. Thus, the watchdog tag would not miss any inquiries made by a RFID reader even if the watchdog tag is further away from the reader than the other tags. Situations where the watchdog tag is not in direct vicinity to the other tags may occur when a user carries its watchdog tag as a separate device or integrated in a cellular phone in her/his pocket and carries a bag with purchased items in a shopping bag.
While the watchdog tag increases a user’s control over the collection of RFID data from tags in the user’s possession, next generation’s RFID tags may offer the capability to deactivate a tag after purchase or will limit direct data access through blocking or encryption. There are three ways to create awareness about RFID tags on products to gain acceptance of these devices among customers. First, use labels to educate consumers on the operation principles and limitations of RFID. Second, enable consumers to control access to the tags on their products. Third, outline possible consumer benefits arising from RFID. People’s fears of privacy invasions are likely to diminish once they are offered direct benefits such as lower prices, faster check-outs, better service, etc. Loyalty cards that usually offer less than 1% of discount have become very popular in recent years, even though loyalty cards provide the retailer with detailed shopping records [149]. While retail loyalty cards are limited to certain chains, mileage programs include not only airlines but also hotels, rental cars, dining, financial services, real estate, and mortgages. The programs provide very detailed information about a person’s shopping and traveling behaviors.

Despite the capability of accessing most RFID tags without authorization, the characteristics of the technology limit random access. Currently used technologies at HF or UHF operate at short distances, which provide minimal scalability for wide-area surveillance. Nevertheless, technology that is embedded in the environment of a retail store such as weight-sensitive foam requires considerably less consumer education and acceptance because the technology’s capability to invade a consumer’s privacy is strongly limited. There may be privacy issues regarding the use of video at a retail store but deploying weight-sensitive foam to monitor shelf stocks suggests no threat to a consumer’s privacy.

### III.4.5 Concluding remarks

Recently, the Grocery Manufacturers Association conducted an industry survey with 31 participants to gain a perceptive on the state of RFID adoption [162]. The report summarizes the progress around case- and pallet-level adoption. Most manufacturers believe that RFID will present long-term value for the industry, but they are skeptical about the integrity of the data for the use in business decision-making. The majority of the participants indicate that retailer pilots failed to generate significant business benefits. Manufacturers also expressed that RFID is not a one-size-fits-all technology and that the benefits and challenges vary significantly by product and business scenario.

Nevertheless, ITechEx estimated the cumulated RFID tag sales in the retail and
consumer goods industry until 2005 at about 230 million tags [163]. The annual sales of RFID tags in 2016 is estimated at 242 billion, which suggest a significant increase of the current market penetration of about 5 percent [4].

Due to the high risks associated with RFID item-level tagging, a combination of RFID at case and pallet level with weight-sensitive foam allows the potential to maximize the advantages of both technologies while minimizing the disadvantages. In this way, the retail industry benefits from RFID through track and trace along the supply chain downstream to the retail store, but avoids the significant costs resulting from equipping each individual product with a tag (most of the benefits derived from RFID at item level can only be achieved if each individual item at the retail store is tagged). Additionally, the customer’s purchase behavior will remain unaltered and no privacy issues will arise.
Table 6: Shows in summary the advantages and disadvantages of RFID and weight-sensitive foam.

<table>
<thead>
<tr>
<th>RFID</th>
<th>Foam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit identification</td>
<td>Inexpensive</td>
</tr>
<tr>
<td>Allows for the detection of misplacements</td>
<td>Stand-alone system</td>
</tr>
<tr>
<td>Supply chain optimization, lower transaction errors, shrinkage reduction, speed-up of checkout</td>
<td>Suggests high adoption rate due to low costs, low risk, and simple installation</td>
</tr>
<tr>
<td>Combination with sensors may provide additional location information</td>
<td>Monitors product arrangement</td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td></td>
</tr>
<tr>
<td>Expensive infrastructure and high variable costs</td>
<td>Non-explicit identification</td>
</tr>
<tr>
<td>Requires broad roll-out and standards for interoperability among supply chain partners</td>
<td>Detection relative to shelf space</td>
</tr>
<tr>
<td>Detection rate (read rates) depends on the product’s material and the product assortment</td>
<td>Limited resolution of product type</td>
</tr>
<tr>
<td>Raises privacy issues</td>
<td>Limited to the application of shelf inventory management</td>
</tr>
<tr>
<td>Local resolution on shelf is limited (inside or outside of the field of the interrogator)</td>
<td>Limited detection of misplacements</td>
</tr>
</tbody>
</table>
The presented automatic shelf stock monitoring technologies suggest improving the replenishment process because they offer sufficient and timely information about shelf stocks. This chapter provides a mathematical analysis of the effect of these technologies on the replenishment process and identifies optimal replenishment strategies based on process costs analysis and customer service level.
IV.1 Inventory Control Systems

IV.1.1 Objectives and approaches

The fundamental reasons for a retailer to carry inventory is to provide a high service level and to display merchandise to the customers at the retail store. On-shelf inventory management requires efficient control of the time for shelf replenishment and the quantity of goods to be restocked. Modern information management and technology create new possibilities for more sophisticated and efficient control of retail shelf inventory. A retailer strives to minimize operation costs in order to increase its profit while maintaining a high service level. Consequently, the value of information derived from inventory monitoring technology is evaluated by its effect on the costs for managing on-shelf inventory with respect to a certain service level.

This thesis makes use of mathematical modeling as a method of operations research to allow comparing the effects of different monitoring technologies on the total replenishment process cost. The mathematical models describe the cost factors of the process and account for the imperfect operation of monitoring technology to support the analysis of average inventory system costs and the development of operating rules for controlling inventory. In order to make the results meaningful and useful, the models developed in this thesis build on general inventory models that are extended to incorporate monitoring technologies.

Despite the fact that operations research methodology is based on mathematical developments that span centuries, it is Ford Harris who is generally credited with the discovery of the original Economic Order Quantity (EOQ) model in 1913, less than a century ago [164]. Significant contributions to the field of mathematical inventory models were made in the early fifties with the AHM-Model, a basic model of inventory with periodic monitoring, formulated by Arrow, Harris, and Marshak [11] and the work on inventory problems with known and unknown distributions of demand by Dvoretzky, Kiefer, and Wolfowitz [165], [166]. Whitin’s “Theory of Inventory Management” in 1953 formed another landmark in the development of mathematical theories of inventories [167]. He was one of the first to describe treatments of the (r,Q) model under uncertainty. These (r,Q) models describe policies where a quantity Q is ordered if the inventory level is smaller or equal to r at a given review point. In 1958, Arrow, Karlin, and Scarf provided a collection of highly sophisticated mathematical models which gave impetus for later work in this area. It was one of the first major analyses of lost sales inventory systems. However,
Hadley and Whitin offer a better introduction and overview of lost sales inventory modeling in their work from 1963 [168]. It was Scarf in 1959 and Iglehart in 1963 who established the optimality of (s,S) policies for single-facility, single-product, periodic review with stochastic demand and full backlogging [169], [170]. The (s,S) models describe order-up-to policies where the order-up-to level is S and the order is triggered if the inventory level is smaller or equal to s at a given review point. In 1969, Morton calculated lower and upper bounds for the optimal inventory level of lost sales model in both finite and infinite horizon problems with linear holding and shortage costs and no fixed order costs [171]. In 1981, Archibald developed a method to calculate optimal values of (s,S) that minimize the long-run average cost when unmet demand is lost in a continuous review problem with constant lead time [172]. Federgruen and Zheng provide efficient methods for calculating optimal (r,Q) policies for continuous review problems and optimal (s,S) policies for periodic review problems with backlogging [173], [174]. In the following, mathematical models have become increasingly complex to account for multi-item, multi-period, multi-echelon problems and for uncertainties in demand, lead time, number of outstanding orders, etc. in a variety of inventory systems. However, many of these efforts exceed the requirements for a sufficient mathematical representation of a retail shelf inventory management system. Indeed, it is critical to find a good abstraction level that limits complexity and provides meaningful results. The mathematical model must be receptive to all relevant influences to the inventory system but neglect factors that deteriorate the effects of the monitoring technology on the inventory system without substantially increasing the accuracy of the model.

IV.1.2 Scope of the descriptive model

The management of on-shelf inventory is concerned with balancing shelf allocation cost (maximum amount of items of a certain product that can be stocked on a shelf), inventory review cost, replenishment cost, costs for the operation of information processing systems, and stock-out cost with respect to customer demand. The on-shelf inventory decreases according to a certain demand pattern and the inventory level is reviewed either continuously or periodically at a constant time interval. The replenishment of the shelf is triggered if the inventory level is detected to be equal to or lower than a certain threshold level at any given review point. The shelf replenishment is completed after a positive lead time, which is constant over time. After replenishment, the inventory level is equal to the base stock level. This means that the shelf is completely restocked regardless of the demand during the replenishment lead time. The assumption of restocking a shelf completely regardless of demand during replenishment contrasts with the general understanding of
inventory systems with positive lead times where inventory levels are raised by the amount ordered at the review point (the time when the inventory level was equal to or smaller than the threshold level). In general, the ordered quantity equals the difference between the base stock and the inventory level at the review point or equals a certain lot size that is identical to or smaller than that difference. However, the order does not account for any demand occurring during the replenishment lead time. Therefore, the inventory level after replenishment is smaller than the base stock level if demand occurs during the replenishment lead time. Consequently, the initial inventory level for a certain period may differ from initial inventory levels in previous periods. Therefore, the inventory management problem needs to be treated as a multi-period problem. However, retail shelves are commonly replenished from boxes and containers up to the base stock level (a sufficiently large reservoir to completely restock the shelf) as opposed to moving only a numbered amount of products from the backroom to the shelf. Super-numerous items are returned to the backroom. This shelf replenishment practice significantly simplifies the on-shelf inventory management problem because each period shows the same initial inventory level. This reduces the complex multi-period problem to a simpler single-period problem without diminishing the significance of the model. The simplification to a single-period problem is only valid under the assumption that the backroom inventory always provides sufficient stocks to completely replenish the shelves. This assumption corresponds to the research problems addressed in this thesis because they are only concerned with on-shelf inventory management and consider backroom inventory management part of a separate effort to optimize store inventory. Therefore, the on-shelf inventory optimization problem becomes independent of inventory holding costs, (they apply in any case) but must account for the cost of allocated shelf space.

With uncertainty in demand and a limited amount of items available on the shelf, there exists a certain probability that customers face stock-out situations. Each stock-out situation is treated as a case of lost sales to the retailer (as opposed to backlogging demand). In order to account for negative effects that are associated with customers facing out-of-stock situations (e.g. dwindling customer loyalty) the model assigns a penalty cost to each stock-out situation in addition to the costs incurred through the loss of sale.

This thesis does not aim to find an optimal replenishment policy for a given technology but uses the widely established threshold-based replenishment policies for periodic and continuous review to analyze and evaluate the effects of shelf stock monitoring technology. With the model’s restrictions described above, the
threshold-based policy is neither a proper \((r, Q)\) nor a proper \((s, S)\) policy. Therefore, the notations \((r, S)\) for the continuous review policy and \((S, r, T_r)\) for the periodic review policy are used where \(r\) is the threshold level, \(S\) the stock-up-to level and \(T_r\) the review period.

In order to calculate long-run average costs with respect to random input variables to the inventory system, it becomes necessary to average over the possible values that the random variables can take on. Expected values represent approximations that become feasible with a sufficiently large planning horizon. A planning horizon of one year suggests adequate approximations and is a meaningful time unit to the retail store manager.

### IV.1.3 Stochastic demand processes

With a retail store selling merchandise to the general public, the predictability of the time patterns of demand is subject to uncertainty. Therefore, demand is described in probabilistic terms, and for the scope of this thesis, the demand pattern is assumed to be generated by a stochastic process. Although demand patterns always change with time (e.g. according to trends, seasonal change, changes due to promotions, etc.), one can assume no change in time if the change is sufficiently slow. This assumption is appropriate especially for items of daily needs.

In this thesis, the occurrence of demand is modeled according to a Poisson process as opposed to a normal distribution to avoid a small probability of negative demand. A Poisson process is defined in terms of occurrences of events. The homogenous Poisson process is characterized by a rate parameter \(\lambda\), such that the number of events in time interval \((t, t + \tau]\) follows a Poisson distribution with associated parameter \(\lambda \cdot \tau\). Consequently, the Poisson process is given as:

\[
p_k(t) = \frac{(\lambda t)^k}{k!} \cdot e^{-\lambda t}, \quad k \in \mathbb{N}_0
\]

where \(e\) is the base of the natural logarithm and \(k!\) the factorial of \(k\). \(\lambda\) is a constant (positive real number) due to the assumption made above that the demand rate does not change over time. Constant demand means that the probability of an instantaneous demand is independent of the time elapsed since the occurrence of the last demand. This “lack of memory” is unique for continuous time considerations; there is no other continuous demand type with this characteristic [175]. This implies that the times between consecutive events are independent random variables and the inter-arrival times are exponentially-distributed with parameter \(\lambda\). \(\lambda\) is also the mean number of occurrences.
IV.1.4 Relevant costs

The costs incurred in operating an inventory system are important when determining an optimal operating policy. The costs that influence the operating policy are only those costs that vary as the policy is changed [168]. Costs independent of the operating policy are not to be included in any analysis where costs are used as an aid in determining an operating policy (idem). The costs considered in this thesis to determine the optimal values of an (r,S) policy are: shelf allocation cost, review cost, replenishment cost, stock-out cost, and cost of operating the information processing system.

A retail store offers limited shelf space to display its products. Shelf allocation cost is an opportunity cost incurred by having a certain amount of shelf space allocated to a product instead of having it allocated to another product. The shelf allocation cost applies for the maximum allocated shelf space and is constant over time. The review cost applies only for the case of manual inventory monitoring where store personnel walk the aisles to visually inspect shelf stocks and to generate picking lists for products that need to be restocked. The replenishment cost incurs each time a shelf has to be physically replenished from the backroom which requires store personnel to move the items from the backroom to the shelf, replenish the shelf from the boxes or cases, and then return super-numerous items to the backroom. The stock-out cost incurs through unmet demand. The cost associated with the sales loss for unmet demand is constant because demand occurs for one unit at a time. Therefore, the cost for sales loss is the loss of profit in not making the sale. Additionally, there is a cost that is associated with the loss of customer loyalty. This may include lost profits on sales of other items or on future sales. The cost of a lost sale also includes special procedures used to inform the customer that a certain demand cannot be met. The cost of operating information processing systems specifically incurs for continuous inventory monitoring with RFID where each product is equipped with a RFID tag. This cost accounts for the cost of the RFID tag which remains on the product past the point of purchase. Although the costs for RFID do not represent a cost for which the inventory management system can be optimized, it is critical to include this cost in order to compare RFID-based inventory management to other approaches such as manual or foam-based inventory management. Consequently, the models should also include infrastructure setup costs with linear amortization to account for the investments made in the infrastructure, e.g. RFID readers and foam.

As mentioned above, inventory holding cost are not considered here. The reason is that the inventory holding cost is not a variable the on-shelf replenishment management system can be optimized for, because backroom inventory is not
included. However, it is assumed that always enough stocks are available in the backroom.

**IV.2 Shelf Inventory Models**

**IV.2.1 Periodic review**

According to subchapter II.2.3, threshold-based periodic review is the most common shelf-stock monitoring practice at retail stores. At the beginning of each periodic interval, the on-shelf inventory positions are inspected and replenishments are triggered if the shelf stocks are equal to or below a certain threshold (Figure 39). If replenishments are triggered, shelves are restocked up to the allocated shelf space after a constant replenishment time $t_l$. The time interval $T_r$ for reviews is one day. This represents common review practices, and therefore, it is not required to treat the time interval as a variable.

Because periodic review is the most common replenishment strategy, the effects of any potential improvements will be compared to the total costs associated with a periodic review inventory management system. Consequently, periodic review forms the benchmark process for improvements.

![Figure 39](image)

**Figure 39:** Shows the inventory level for periodic review and replenishment with different lead times.

Replenishment is triggered if, and only if, the inventory level at a review time is equal to or smaller than the threshold. The time between two replenishments is an integral multiple of the time $T_r$ plus lead time $t_l$. However, the number of periods included in a cycle is a random variable. Consequently, the length of a cycle is also
a random variable. In order to derive the cycle time $T_c$ for periodic review, which is:

$$E[T_c] = E[n] \cdot T_r + t_i$$  \hspace{1cm} \text{Eq. (7)}$$

the number of reviews per cycle needs to be determined. An extensive analysis of periodic review inventory systems was done by Hadley and Whitin [168]. Their findings are adapted for the analysis of a periodic review systems considered in this thesis. A cycle will be exactly one period if the demand in the first period is greater than $S-r$. The probability of this is $P(S-r; T_r)$, where $P(x; T_r)$ is the complementary cumulative of $p(x; T_r)$. A cycle will contain precisely $n$ ($n \geq 2$) periods if after a demand of $S-r-j$ items in the first $n-1$ periods after replenishment, $j$ or more items are demanded in period $n$. Thus, the probability that a cycle contains precisely $n$ periods is:

$$\sum_{j=1}^{S-r} p^{(n-1)}(S-r-j; T_r) \cdot P(j; T_r)$$  \hspace{1cm} \text{Eq. (8)}$$

where $p^{(n-1)}(x; T_r)$ is the $(n-1)$-fold convolution of $p(x; T_r)$, $S$ the restock-up-to level, and $r$ the threshold. Hence the expected number of periods in a cycle is:

$$E[n] = P(S-r; T_r) + \sum_{n=2}^{\infty} \sum_{j=1}^{S-r} n \cdot p^{(n-1)}(S-r-j; T_r) \cdot P(j; T_r)$$  \hspace{1cm} \text{Eq. (9)}$$

$$= \sum_{n=1}^{\infty} \sum_{j=1}^{S-r} n \cdot p^{(n-1)}(S-r-j; T_r) \cdot P(j; T_r)$$  \hspace{1cm} \text{Eq. (10)}$$

In the case of Poisson demand, $p(x; T_r)$ becomes $p(x; \lambda \cdot T_r)$ and $p^{(n)}(x; T_r)$ becomes $P(x; n \cdot \lambda \cdot T_r)$.

Hence, the equation for the expected number of periods in a cycle (Eq. (10)) becomes:

$$E[n] = \sum_{n=1}^{\infty} \sum_{j=1}^{S-r} n \cdot p[S-r-j; (n-1) \cdot \lambda \cdot T_r] \cdot P[j; \lambda \cdot T_r]$$  \hspace{1cm} \text{Eq. (11)}$$

With

$$p[S-r-j; (n-1) \cdot \lambda \cdot T_r] = e^{-(n-1) \cdot \lambda \cdot T_r} \cdot \frac{(\lambda \cdot T_r \cdot (n-1))^{S-r-j}}{(S-r-j)!}$$  \hspace{1cm} \text{Eq. (12)}$$

which is the probability that the demand in $(n-1) \cdot T_r$ is equal to $S-r-j$, and

$$P[j; \lambda \cdot T_r] = \sum_{x=j}^{\infty} e^{-\lambda \cdot T_r} \cdot \frac{(\lambda \cdot T_r)^x}{x!}$$  \hspace{1cm} \text{Eq. (13)}$$
which is the probability that the demand in $T_r$ is equal to or greater than $j$, the expected number of periods in a cycle becomes:

$$E[n] = \sum_{n=1}^{\infty} \sum_{j=1}^{S-n} n \cdot e^{-\lambda \cdot T_r \cdot (n-1) \cdot T_r} \cdot \left( \frac{(\lambda \cdot T_r \cdot (n-1))^{S-r-j}}{(S-r-j)!} \right) \cdot \sum_{x=j}^{\infty} e^{-T_r \cdot \lambda \cdot T_r \cdot x} \cdot \left( \frac{\lambda \cdot T_r \cdot x}{x!} \right)$$

Eq. (14)

Note, if the demand in $T_r$ is equal to or greater than $j$, the inventory level is equal to or smaller $r$ and replenishment is triggered.

Using the relation in Eq. (15),

$$\sum_{n=0}^{\infty} \frac{(x)^n}{n!} = e^x$$

Eq. (15)

Eq. (14) is transformed to:

$$E[n] = \sum_{n=1}^{\infty} \sum_{j=1}^{S-n} n \cdot e^{-\lambda \cdot T_r \cdot (n-1) \cdot T_r} \cdot \left( \frac{(\lambda \cdot T_r \cdot (n-1))^{S-r-j}}{(S-r-j)!} \right) \cdot \left( 1 - e^{-\lambda T_r} \cdot \sum_{m=0}^{j-1} \frac{(\lambda T_r)^m}{m!} \right)$$

Eq. (16)

According to Eq. (7), the cycle time is now derived as $E[n] \cdot T_r + t_i$. The number of replenishments per planning horizon $T$ is simply $T$ divided by the cycle time.

The on-shelf inventory management system is modeled as a single-period system. This system repeats itself after each cycle with cycle of length $T_c$. The single-period cost function $C_{Tc}(r, S, T_r)$ is then formulated as:

$$C_{Tc}(r, S, T_r) = c_s \cdot S \cdot \frac{E[T_c]}{T} + E[n] \cdot c_r + c_{rep} + p \cdot E[y^-]$$

Eq. (17)

where $c_s$ is the shelf allocation cost per item per year, $c_r$ the review cost of a store clerk for visually inspecting on-shelf inventory positions, $c_{rep}$ the replenishment cost, $p$ the penalty for shortages, $E[y^-]$ the expected number of shortages per cycle, and $y$ the inventory level. Shortages occur if the demand in the $n$-th period plus lead time is greater than the on-hand inventory at the end of the $(n-1)$-th period. The expected number of shortages is given by:
The last term is the expected value for demand greater \( r + j \) given \( j \), and the previous term is the probability that the inventory positions is at \( S - r - j \) after \( (n - 1) \) periods. The expected number of shortages is:

\[
E[y^-] = \sum_{n=0}^{\infty} \sum_{j=1}^{S-r} \left( \sum_{x=r+j}^{\infty} (x-r-j) \cdot p(x; \lambda(T_r + t_l)) \right) \cdot \left( \sum_{x=r+j}^{\infty} (x-r-j) \cdot e^{-\lambda(T_r + t_l)} \cdot \frac{(\lambda(T_r + t_l))^x}{x!} \right)
\]

Eq. (19)

According to [168],

\[
\sum_{j=r}^{\infty} (j-r) \cdot p(j; \lambda T_r) = \sum_{j=r}^{\infty} j \cdot p(j; \lambda T_r) - r \cdot P(r; \lambda T_r)
\]

Eq. (20)

\[
= \lambda T_r \cdot P(r-1; \lambda T_r) - r \cdot P(r; \lambda T_r)
\]

Eq. (19) can be transformed to:

\[
E[y^-] = \sum_{n=0}^{\infty} \sum_{j=1}^{S-r} \left( \sum_{x=r+j}^{\infty} (x-r-j) \cdot e^{-\lambda(T_r + t_l)} \cdot \frac{(\lambda(T_r + t_l))^x}{x!} \right)
\]

Eq. (21)

in order to avoid one summation up to infinity.

The customer service level represents the availability of products and relates to the out-of-stock rate. In general, the stock-out rate is given by the amount of products that are out-of-stock in relation to the total number of SKUs carried by a retailer at
any given time. For the analysis of single-item inventory management, the stock-out
rate is defined as the number of unmet demands or number of units short in relation
to total demand per replenishment cycle. An unmet demand occurs when an arriving
customer faces an out-of-stock situation. A depleted shelf without the occurrence of
demand does not affect the customer service level. The customer service level $S_L$ is
given as:

$$S_L = 1 - \left( \frac{E[y^-]}{\lambda \cdot E[T_c]} \right)$$

Eq. (22)

The customer service level is not directly reflected by the single-period cost
function but its effect contributes to the system costs through a penalty cost for
unmet demand.

The long-run average cost is calculated for each set of $(r,S)$ control parameters
($T_r = 1$). The set that minimizes the cost function represents the control parameters
of the optimal $(r,S,T_r)$ policy. The long-run average cost function is given by:

$$C(r, S, T_r) = \left( c_s \cdot S \cdot \frac{E[T_c]}{T} + E[n] \cdot c_r + c_{rep} + p \cdot E[y^-] \right).$$

$$\frac{T}{E[T_c]}$$

Eq. (23)

IV.2.2 Continuous review

In contrast to the previous model that relies on periodic review, the model
developed in this subchapter assumes continuous review. Replenishment is triggered
when the inventory position has declined to the threshold level $r$ according to
Poisson demand with demand size one (Figure 40). After the replenishment lead
time, the inventory level is raised up to $S$. In contrast to periodic review, the major
advantage of continuous review is that the same level of customer service is
achieved with less safety stock [176]. This outcome results from the fact that the
period during which safety protection is required, is longer under periodic review
(idem). On the other hand, continuous review is considered more expensive than
periodic review. However, if the review costs are sufficiently low, continuous
review may result in lower inventory management cost than periodic review. The
model developed in the following assumes continuous review with no review costs.
The effects of monitoring technology costs on the inventory management system
and the uncertainties in detections are addressed later.
Customers arrive according to a Poisson process and S-r items are sold until replenishment is triggered (Figure 40). The time from the beginning of a cycle until replenishment is triggered is the cumulative time for S-r inter-arrivals. For Poisson demand, the expected value for the time \( t_r \) after which replenishment is triggered is:

\[
E[t_r] = \frac{S-r}{\lambda}
\]

Eq. (24)

and consequently, the expected value for an entire cycle time is:

\[
E[T_c] = E[t_r] + t_l
\]

Eq. (25)

with cycle time \( T_c \) and replenishment lead time \( t_l \).

The on-shelf inventory management system is modeled as a single-period system that repeats itself after a cycle of length \( T_c \). The single-period cost function \( C_{T_c}(r,S) \) is given by:

\[
C_{T_c}(r,S) = \frac{K_I \cdot S \cdot E[T_c]}{T} + \frac{c_s \cdot S \cdot E[T_c]}{T} + c_{rep} + p \cdot E[y^-]
\]

Eq. (26)

where \( K_I \) is the infrastructure cost per allocated shelf space per item per year, \( c_s \) the shelf allocation cost per item per year, \( c_{rep} \) the replenishment cost, \( p \) the penalty for shortages, \( E[y^-] \) the expected number of shortages per cycle, and \( y \) the inventory level. Shortages only occur if the demand during lead time is greater than \( r \). Consequently, the expected number of shortages is:
\[ E[y^-] = \sum_{k = r + 1}^{\infty} k \cdot e^{-\lambda t_i} \cdot \left( \frac{(\lambda t_i)^k}{k!} \right) \tag{27} \]

Solving this equation requires extensive computing because of a summation up to infinity, the factorials of very large numbers, and \((\lambda \cdot t_i)\) to the power of very large numbers. Using the relation in Eq. (15), Eq. (27) is transformed to:

\[ E[y^-] = \sum_{k = r + 1}^{\infty} k \cdot e^{-\lambda t_i} \cdot \left( \frac{(\lambda t_i)^k}{k!} \right) = \lambda \cdot t_i \cdot \left( 1 - e^{-\lambda t_i} \cdot \sum_{m = 0}^{r-1} m \cdot \frac{(\lambda t_i)^m}{m!} \right) \tag{28} \]

The customer service level \(S_L\) is again given as:

\[ S_L = 1 - \left( \frac{E[y^-]}{\lambda \cdot E[T_c]} \right) \tag{29} \]

The long-run average cost function is a multiple of the single-period cost function. The multiplier is the ratio between the planning horizon and the expected value of the replenishment cycle time. By making use of expected values, the single-period cost function represents an approximation. However, with a sufficiently large planning horizon the long-run average cost function represents a satisfactorily approximation. A planning horizon of one year suggests being adequate for mathematical considerations and suitable for economical analysis. The long-run average cost function is therefore given as:

\[ C(r, S) = K_I \cdot S + c_S \cdot S + c_{rep} \cdot \frac{T}{E[T_c]} + p \cdot E[y^-] \cdot \frac{T}{E[T_c]} \tag{30} \]

The optimal \((r,S)\) policy is given by those parameters \(r\) and \(S\) that minimize the long-run average cost function. The computation of optimal parameters for the long-run average cost function requires limited processing power due to the use of expected values. The cost function is evaluated for each possible set of \(r\) and \(S\). The number of evaluations of the cost function is of the order \(n \cdot (n-1)\) where \(n\) represents the upper limit for \(S\) that is the maximum number of items for which shelf space may be allocated.

**IV.3 Shelf Inventory Models with Imperfect State Information**

Radio frequency identification has gained a lot of attention in white papers and research reports that examine potential benefits arising from applying RFID to supply chain logistics (cf. III.2.2). However, many of these papers and reports fail to thoroughly analyze how the introduction of RFID influences and changes the
individual logistics processes and how the actual benefit from RFID is derived.

Among the first to examine the influence of RFID on the fundamental operating characteristics of a system based on analytical models are Kang [177], Kang & Gershwin [182], Gaukler [178], and Lee et al. [63]. Lee et al. examine the impact of shrinkage, misplacements, and transaction errors to evaluate the potential value of RFID. They show how actual inventory and recorded inventory may diverge over time due to errors that are introduced to the system such as shrinkage, misplacements, and transaction errors and the implications for inventory management thereof based on total process costs. In addition, they apply the visibility characteristics of RFID to their first model and evaluate the impact on the system to constitute the value of RFID. They accredit RFID two distinct values – visibility and prevention. Visibility allows the store manager to identify errors in the inventory record. Isolating these errors serves to prevent redundant operations and reduce costs. By comparing their RFID-enabled model with models that lack visibility, they derive explicit numbers for the reduction of total cost due to RFID.

Gaukler examines the benefits of item-level RFID for a retail setting with a focus on product availability on the retail shelf. In his model, Gaukler assumes that the backroom stock is replenished once at the beginning of a selling season. The shelf on the sales floor that has limited shelf space is replenished from the backroom stock on a frequent basis. This model allows evaluating the effectiveness of RFID for the shelf replenishment process in the case of shrinkage, misplacements, and other execution errors. It also allows deriving the break-even tag price, the price at which one is indifferent between RFID and no RFID for any given set of model parameters. While Gaukler and Lee et al. make use of RFID in order to increase visibility and to eliminate transaction errors, their models do not account for imperfect state information due to the physical characteristics of the technology. Such characteristics may be read rates of less than 100% due to tag shielding or due to failures of tags. In addition, there may be failures in operating the reader or processing the RFID data. Kang and Kang & Gershwin seem to be the first ones to specifically address inventory control systems with imperfect state information due to limited read rates of RFID tags. While they identified the general causes (e.g. shrinkage, misplacements, and transaction errors) for information inaccuracy in inventory systems and analyzed the benefits for inventory management derived from RFID, they also account for the fact that observations are subject to measurement errors. Kang provides an approach that treats this inventory control problem as an imperfect state information problem where the measurement data represents stochastically uncertain observations of the stock quantities. Although imperfect state information dynamic programming, as suggested by Kang, provides a means to
determine the optimal policy for a problem that is subject to measurement errors, it requires highly intensive computing as the dimension of the information vector grows excessively with time. The information vector contains all information available at a certain time, specifically all previous observations and all past controls. In addition, the information vector is used instead of the actual state of the system because the actual state is not accessible. The fast-growing state dimensions of the imperfect state information problem limits the feasibility of this approach to small and simplified problems. Kang provides an optimal policy for a control system with a planning horizon of only 5 days, whereas a planning horizon of 365 days is more desirable.

Other research on shelf replenishment has made use of simulations to investigate the effects of inventory inaccuracy on shelf replenishment. Lee et al. present a simple simulation model of a three echelon supply chain in order to examine the effect of RFID deployment on the supply chain [179]. They studied the improvement of inventory management due to RFID if the inventory is subject to shrinkage. They also examined the effect of RFID on the shelf replenishment process where the on-shelf inventory level is updated continuously (e.g. through RFID readers in the shelves), and they compare this approach to the traditional replenishment process based on periodic review. They argue that a RFID-controlled replenishment process excels the periodic review process because it removes the uncertainty of remaining shelf stock quantities. Therefore, a RFID-controlled replenishment process allows for timely replenishments and lower overall safety stocks. However, the significance of their findings is limited because they chose arbitrary inventory policies in order to examine the performance of the different replenishment strategies instead of optimized policies. Additionally, the authors consider only stock levels and stock-outs as the main performance indicator instead of total cost. Finally, Lee et al. do not account for any failures in capturing relevant RFID data.

Thiesse et al. examine a retail store setup with a gate-reader between the backroom and the sales floor to keep track of merchandise that is moved onto the sales floor. They present a simulation study on RFID at case-level and its effect on the shelf replenishment process [180]. They also present the equivalent study for RFID at item-level [181]. In both simulations, they assume that the RFID tags are detected at a certain read rate when moved from the backroom onto the sales floor. For case-level tagging, the sales floor inventory is increased by the number of units per box each time the movement of a box from the backroom onto the sales floor is detected. Thiesse et al. account for the imperfect operation of RFID by including a read rate at which the boxes are detected. The sales floor inventory is reduced each time an item
is sold and its barcode is scanned at the check-out counter. However, their model does not account for the significant data inaccuracies introduced through irregularities at the check-out (cf. II.3.3). This omission constitutes a considerable distortion of the estimated inventory level and diminishes the positives effects of RFID on inventory visibility. After shelf replenishment, the empty case is moved to the trash compactor where the RFID tag on each case is read. They argue that this process allows detecting cases that were not identified previously when moved from the backroom to the sales floor and that the sales floor inventory level could be adjusted a posteriori. Their model does not account for imperfect operation of RFID at the trash compactor where read rates are assumed 100% or for the return of cases to the backroom that have not been emptied completely. They assume that all units of a case fit into the available shelf space when a shelf is replenished. However, they do not account for this restriction when optimizing for maximum allocated shelf space. In addition, they do not address how the optimal values for a system are affected if the inventory level is increased by multiples of units in a case that results in an inventory level smaller than S.

If cases are not detected at either location, the trash compactor and the gate between backroom and sales floor, estimated sales floor inventory levels will be significantly distorted. The reason is that the actual inventory level will be higher than the recorded inventory by the number of units in a box. Additionally, their model does not account for any delays that may occur between restocking the shelf and detecting the case’s tag at the trash compactor. Thiesse et al. apply the equivalent procedure for sales floor inventory estimations for the case of RFID item-level tagging. In this case, the read rate at the gate between the backroom and the sales floor applies to each individual unit. Again, the recorded sales floor inventory is reduced whenever an item is sold. In this case, the data generated at the check-out can be assumed accurate because it relies on RFID. However, their model for RFID at item level does not account for the stochastic transient nature of the recorded inventory that occurs when items are on the sales floor that were not detected when moved there. Moreover, the recorded inventory level is not completely adjusted until all items are sold that were not detected previously. If those items are sold first, the estimated sales floor inventory is assumed low and replenishment may be triggered even though the actual inventory may still be high. The discrepancy between the recorded inventory and the actual inventory is not adjusted until all previously undetected items are sold.

While their approach of estimating the inventory level based on a limited number of read points (such as the gate reader between the backroom and the sales floor as well as at the check-out) is attractive due to low infrastructure cost, the information
on shelf availability derived from this data is still subject to uncertainty. Once a case or several units have been moved onto the sales floor, they may or may not be stocked on a certain shelf, and significant delays may occur between moving the items onto the sales floor and the actual restocking.

In fact, if the shelf is always refilled up to S, as they assume in the case of RFID item-level tagging, the information about read rates is useless. Instead, the information that replenishment has occurred is the only relevant information. Capturing a replenishment event requires the detection of moving merchandise on the sales floor and is independent of the amount. Simply detecting the event of replenishment could certainly be performed with high accuracy. Detecting the event of replenishment would reset the recorded inventory level to S, and each time an item is detected at the check-out, the recorded inventory level is reduced by one.

In the following two subsections, inventory models are developed to analyze the impact of on-shelf inventory monitoring technology on the total replenishment cost. The models account for measurement errors specific to RFID and weight-sensitive foam. The models examine the progression of the recorded inventory over time and its implication for replenishment and total process costs. RFID and weight-sensitive foam are applied for shelf stock monitoring as described in Section III. In the case of RFID item-level tagging, a RFID reader antenna is incorporated into the shelf to continuously read tags (equivalent to a read event each time a demand occurs). The tags remain on the products past the point of sale and are not recyclable. Weight-sensitive foam forms an intermediate layer between the shelf and the products on display. The products are detected through their gravity force that is applied to the foam. The foam remains on the shelf when the products are removed.

**IV.3.1 Continuous review with imperfect state information based on RFID**

The model developed in this subsection shows the same basic characteristics as the previous one. The model assumes a customer arrival pattern according to a Poisson process. Replenishment is triggered exactly when the recorded inventory position reaches threshold r and the on-shelf inventory level is raised up to S after lead time $t_l$. The difference between this model and the previous model is constituted by the recorded inventory that may diverge from the actual inventory (Figure 41). A difference in recorded and actual inventory is due to measurement errors caused by the reading of RFID tags. The inventory monitoring is performed by RFID antennae integrated into the shelves. The antennae detect the tags attached to items on display.

While there are many reasons why an item may not be detectable, such as a
malfunctioning or broken tag, a tag that has been removed, interoperability issues, etc., the only cause for a tag to not be readable considered in this model is shielding. The shielding that the model accounts for is caused by another item in close proximity that either prevents the RFID signal from penetrating or that detunes the RFID tag temporarily, thus it would not properly respond to RFID interrogator requests. Consequently, only a certain amount of the products on display are detected by the RFID reader. The reader introduces a measurement or detection error into the inventory management system that results in a difference between recorded and actual inventory. However, the inventory management system incorporates a refill-up-to S policy, which means that the actual inventory level at the beginning of each cycle is S. By using this information and combining it with the result of the initial inventory measurement, the difference of the actual and the recorded inventory at the beginning of each cycle is determined. Replenishment is triggered when the recorded inventory level reaches the threshold but penalties for stock-outs occur only if the actual inventory does not meet demand. In order to determine optimal values for an (r,S) policy with respect to lowest cost, it is important to analyze the progression of detection errors over an entire replenishment cycle.

Figure 41: Illustrates the progression of actual inventory and recorded inventory over time and the measurement error that constitutes the difference between them. The upper line represents the actual inventory level and the lower one the recorded inventory level. At $k=2$, $4$, and $5$, a previously invisible item becomes visible while an item is removed. This reduces the measurement error by one for each occurrence.
The initial difference between actual and recorded inventory at the beginning of a new cycle is defined as $\Delta_0$. The initial difference is equal or greater zero depending on the initial read rate. The recorded inventory is always equal or smaller than the actual inventory because measurement errors only occur when a tag is not detected. The opposite error, that a tag is detected that does not exist, is not possible due to the operating principles of RFID. The recorded inventory shows a high probability for being smaller than the actual inventory. Therefore, the inventory system is likely to trigger replenishments early which results in more frequent replenishments than necessary. However, the expected time for replenishment depends on the progression of the difference between the actual inventory and the recorded inventory over time. It was initially stated that measurement errors only occur through shielding caused by other items on display. Assuming that only one item is left after S-1 customers have arrived, the one remaining item cannot be shielded by any other items. Therefore, this tag must be readable. Measurement errors in the inventory management system must have decreased to zero by this time. The measurement error decreases from an initial value of $\Delta_0$ to zero over S-1 steps according to a stochastic process unless replenishment occurs.

In order to limit the complexity of measurement errors’ progression over time, the maximum number of items that can be shielded by another item is restricted to one (Figure 41). Although this constitutes a loss of generality, it still suggests feasibility for a profound analysis of the system’s behavior and for meaningful results. Hence, if an item is removed it may or may not reveal another item that becomes visible to the RFID system. The number of items revealed in one cycle cannot exceed the number of measurement errors recorded at the beginning of that cycle. The inventory management system only has knowledge about the inventory level of the recorded inventory that is subject to measurement errors and the initial value of the actual inventory. However, recorded inventory and actual inventory conform over time. For each event of demand the actual inventory is reduced by one. If the item that is removed from the shelf has not revealed another item that now becomes visible to the system, the recorded inventory is also reduced by one and the difference between the two inventory counts remains the same. If, however, an item does become visible due to the fact that another item has been removed, the recorded inventory level remains at the same level, while the actual inventory is reduced by one. Consequently, the difference between the two inventory levels is reduced by one as well. The occurrence of an item becoming visible to the system when demand occurs is modeled as a stochastic process. This stochastic process depends on the state of the system and the difference between the two inventory counts.
The progression of the recorded inventory level \( y_{rk} \) is described as a function of the actual inventory.

\[
y_{rk} = S - \Delta_0 - k + \xi_k
\]  

Eq. (31)

where \( S \) is the initial value of the actual inventory, \( \Delta_0 \) the initial measurement error, \( k \) the number of demands that have occurred since the beginning of a cycle, and \( \xi_k \) is the cumulative of all items at state \( k \) that have become visible since the beginning of a cycle. \( \xi_k \) is given by:

\[
\xi_k = \sum_{i=0}^{k} \varepsilon_i, \quad \varepsilon_i \in \{0, 1\}
\]  

Eq. (32)

where \( \varepsilon_i \) is the event that at state \( i \) an item becomes visible (\( i \leq k \)). The states, at which items become visible during a period of \( S-1 \) events of demand, show a binominal distribution. There are \( \Delta_0 \) items that eventually become visible during the \( S-1 \) events of demand, and therefore, there are \( S-1-\Delta_0 \) events of demand at which an item is removed without another one becoming visible.

Replenishment occurs when \( y_{rk} \) is equal to the threshold \( r \) which is equivalent to:

\[
\xi_k = r - S + \Delta_0 + k
\]  

Eq. (33)

This is simply the transformation of \( y_{rk} = r \), thus \( \xi_k \) is on one side of the equation and the remaining terms on the other side.

Let \( P(\xi_k = n) \) be the event that in a sequence of \( k \) occurrences of demand, \( n \) items have become visible. Then:

\[
P(\xi_k = n) = \frac{[\xi_k = n]}{[\Omega]}
\]  

Eq. (34)

which is the number of results in \( \xi_k \) divided by the number of all possible results. \( [\Omega] \) is given by:

\[
[\Omega] = (S-1)_k = \frac{(S-1)!}{(S-1-k)!}
\]  

Eq. (35)

where “!” is the factorial of the term in parentheses. The sequence of \( n \) events of an item becoming visible over a period of \( k \) steps can be arranged in \( \binom{k}{n} \) possible ways. For each of these \( \binom{k}{n} \) different positions, there exist \( (\Delta_0)_n \) \((S-1-\Delta_0)_{(k-n)}\) \( k \)-tuples.
Therefore,
\[
P(ξ_k = n) = \binom{k}{n} \cdot \frac{\Delta_0 (S-1-\Delta_0)}{(S-1)_k}
\]
Eq. (36)

\[
= \binom{n}{k} \frac{(S-1-\Delta_0)}{k-n}
\]
Eq. (37)

(cf. B.1 for a proof of Eq. (37))

Based on the probability distribution for ξ_k, the expected time until replenishment is calculated as the cumulative for all states S and the time to go from the initial state S to state S-k multiplied by the probability that the recorded inventory at this state has just dropped down to the threshold level. Note that the recorded inventory level may remain at the same level despite the occurrence of demand because items may become visible. Therefore, it is important to only account for the time elapsed until the recorded inventory level drops down to r for the first time. This is the time after which the replenishment is triggered.

The expected time until replenishment is given by:

\[
E[t_r] = \sum_{k=1}^{S} \frac{k}{\lambda} \cdot prob\left( S-\Delta_0-k+ξ_k = r \mid S-\Delta_0-(k-1) \right)
\]
\[
+ prob\left( S-\Delta_0-(k-1)+ξ_{k-1} = r+1 \mid S-(k-1) \right)
\]
Eq. (38)

where the first term is the time to go from S to S-k, and the second term is the probability that the recorded inventory level is at r+1 at state k-1 (just one above the threshold level) and that the recorded inventory level drops down to r when the next demand occurs (which assumes that no item becomes visible with the latest event of demand). Applying Bayes’ theorem of \( P(A \cap B) = P(A|B) \cdot P(B) \), the expected time until replenishment is reformulated as:

\[
E[t_r] = \sum_{k=1}^{S} \frac{k}{\lambda} \cdot prob\left( S-\Delta_0-k+ξ_k = r \mid S-\Delta_0-(k-1) + ξ_{k-1} = r+1 \right)
\]
\[
\cdot prob\left( S-\Delta_0-(k-1)+ξ_{k-1} = r+1 \mid S-(k-1) \right)
\]
Eq. (39)

Computing the probability that the inventory level drops from r+1 at state k-1 to r at state k is simple. It is the probability that with the next demand no item becomes
visible given that the system is at state k-1. If the system is at state k-1, there are S-1-(k-1) states remaining until all shielded items have become visible. At state k-1, \( \xi_{k-1} \) out of \( \Delta_0 \) have already become visible which leaves \( \Delta_0 - \xi_{k-1} \) items that have yet to become visible during the remaining S-1-(k-1) steps. Inversely, during these S-1-(k-1) steps, the number of times an item does not become visible is S-1-(k-1)-(\( \Delta_0 - \xi_{k-1} \)). Substituting \( \xi_{k-1} \) with \( r+1-S+\Delta_0+(k-1) \) results in the number of times an item does not become visible which is equal to \( r \). Exactly one of these occurrences is required, thus the recorded inventory level will drop from \( r+1 \) to \( r \) with the next demand. Consequently, the probability of this occurrence is:

\[
 prob(\xi_k = 0) = \frac{r}{S-1-(k-1)} = \frac{r}{S-k} \tag{40}
\]

The probability that the inventory level is at \( r+1 \) at state k-1 is transformed as follows:

\[
 prob(S - \Delta_0 - (k-1) + \xi_{k-1} = r + 1 | S - (k-1)) = prob(S - \Delta_0 - k + \xi_{k-1} = r | S - k + 1)
\]

\[
 = prob(\xi_{k-1} = r - S + \Delta_0 + k | S - k + 1) \tag{41}
\]

The probability \( P(\xi_k = n) \) for \( \xi \) at state k was given in Eq. (37). For \( \xi \) at state k-1, \( P(\xi_k = n) \) transforms to:

\[
P\left(\xi_{k-1} = n\right) = \frac{\binom{\Delta_0}{n} \binom{S-1-\Delta_0}{k-n}}{\binom{S-1}{k-1}} \tag{42}
\]

with \( n = r-S+\Delta_0+k \).

Hence, the expected time until replenishment is:

\[
 E[t_r] = \sum_{k=1}^{S} \frac{k}{\lambda} \cdot \frac{r}{S-k} \cdot \frac{\binom{\Delta_0}{n} \binom{S-1-\Delta_0}{k-n}}{\binom{S-1}{k-1}} \tag{43}
\]

where \( t_r \) is the time from the beginning of a cycle until replenishment. The expected cycle time is:

\[
 E[T_c] = E[t_r] + t_l \tag{44}
\]

The number of shortages depends on the demand during lead time and the actual
The actual inventory level at time $t_r$ is greater or equal to $r$ depending on the difference between the actual and the recorded inventory at that time. Shortages occur if the demand during lead time exceeds the available inventory. The expected number of shortages for a cycle is the expected number of shortages at state $k$ multiplied by the probability that the recorded inventory level has just dropped to $r$ and computed for all $k$. Consequently, the expected number of units short is given by:

$$E[y^-] = \sum_{k=1}^{S-1} \sum_{x=S-k}^{\infty} (x-(S-k)) \cdot e^{-\lambda t_l} \cdot \left(\frac{\lambda t_l}{x!}\right)^x.$$  

Eq. (45)

Making use of the findings in Eq. (40) and Eq. (42), Eq. (45) becomes:

$$E[y^-] = \sum_{k=1}^{S-1} \left[ \frac{r}{S-k} \cdot \left(\frac{\Delta_0}{n}\right) \cdot \left(\frac{S-1-\Delta_0}{k-n}\right) \cdot e^{-\lambda t_l} \cdot \sum_{x=S-k}^{\infty} (x-S+k) \cdot \left(\frac{\lambda t_l}{x!}\right)^x \right].$$  

Eq. (46)

The customer service level $S_L$ is given by:

$$S_L = 1 - \left(\frac{E[y^-]}{\lambda \cdot E[T_c]}\right).$$  

Eq. (47)

The long-average cost function is formulated as:

$$C(r, S)_{\phi} = \frac{T}{E[T_c]} \cdot \left(\frac{K_l \cdot S \cdot E[T_c]}{T} + \frac{c_s \cdot S \cdot E[T_c]}{T} + c_{rep} + p \cdot E[y^-] + c_{tag} \cdot \min\left(S, \lambda \cdot E[T_c]\right)\right).$$  

Eq. (48)

where $\phi$ indicates the mean initial read rate $(S-\Delta_0)/S$, $T$ the planning horizon, $K_l$ the infrastructure cost per allocated shelf space per item per year, $c_s$ the shelf allocation cost per item per year, $c_{rep}$ the replenishment cost, $p$ the penalty for shortages, $E[y^-]$ the expected number of shortages per cycle, and $y$ the inventory level, and $c_{tag}$ the
tag cost for each item sold (satisfied demand). The initial read rate is a stochastic variable; however, it can be approximated with an expected value corresponding to the results of numerous measurements.

The recorded inventory has a high probability of showing a smaller inventory level than the actual inventory which results in early replenishment. Although the expected long-run average cost function already accounts for the replenishment cost, it is of interest to analyze the additional number of replenishment per planning horizon for illustrative purposes. Let \( N_{r1} \) be the number of replenishments if there were no measurement errors. Then,

\[
N_{r1} = \frac{T}{E[t_{r1}]}, \quad E[t_{r1}] = \frac{S-r}{\lambda}
\]

Eq. (49)

Let \( N_{r2} \) be the number of replenishments in the presence of measurement errors. Then,

\[
N_{r2} = \frac{T}{E[t_{r2}]}
\]

Eq. (50)

while \( E[t_{r2}] \) is the expected time until replenishment that was determined for a system that is subject to measurement errors. Consequently, the additional number of replenishments per planning horizon is the difference \( N_{r2} - N_{r1} \)

\[
N_{r1} - N_{r2} = T \cdot \left( \frac{1}{E[t_{r2}]} - \frac{1}{E[t_{r1}]} \right)
\]

Eq. (51)
IV.3.2 Continuous review with imperfect state information based on weight-sensitive foam

In order to compare the effects on inventory management for continuous review based on RFID and weight-sensitive foam, the models show the same basic characteristics. While weight-sensitive foam has been developed to provide relative information about stock quantities with regard to loaded shelf space, its detection mechanism has to be modified to provide absolute quantity information in order for it to be comparable to RFID. Recent research has developed an algorithm to extract absolute product quantities from the pressure pattern of items on display generated by the weight-sensitive foam presented in Section III.3 [136]. While measurement errors for an inventory system based on RFID occur through shielded tags, the detection of product quantities with weight-sensitive foam is not susceptible to other products. However, measurements may occur if loaded sensors do not report a load and if unloaded sensors report a load due to poor product alignment (cf. III.3). In contrast to measurement errors that may occur with RFID where the recorded inventory level is equal to or smaller than the actual inventory level, weight-sensitive foam may introduce false negatives as well as false positives (the false detection of items and undetected items). While false negatives reduce the recorded inventory, false positives increase it. Therefore, the inventory system has to account for both types of measurement errors as well as their superposition.

For continuous review based on RFID, the complexity of the analysis of the measurement errors’ progression over time was reduced by introducing the restriction that each item cannot shield more than one other item and that if only one item is left on the shelf, the measurement error has decreased to zero. In the case of continuous review based on weight-sensitive foam, each measurement error is caused by exactly one item (for false negatives: an item covers a sensor that does not respond to the load; for false positives: an item causes an unloaded sensor to report a load due to poor alignment). Consequently, the assumption that one or both measurement errors may be reduced by one when an item is removed from the shelf does not constitute a loss of generality. Similar to the case with RFID, once all items are removed from the shelf both measurement errors will have decreased to zero (it was shown in Section III.3 that weight-sensitive foam is highly robust to false positives in an unloaded state). False negatives and false positives are treated as two separate measurement errors and the occurrences of measurement error reductions are treated as two independent stochastic processes. The progression of the recorded inventory level $y_{rik}$ is described as a function of the actual inventory analogous to Eq. (31).
\[ y_r(k) = S - (\Delta p_0 + \Delta n_0) - k + (-\xi p_k + \xi n_k) \]  

Eq. (52)

where \( S \) is the initial value of the actual inventory, \( \Delta p_0 \) the initial measurement error for false positive, \( \Delta n_0 \) the initial measurement error for false negatives, \( k \) the number of demands that have occurred since the beginning of the cycle, and \( \xi p_k \) the cumulative of the reduction in false positive at state \( k \), and \( \xi n_k \) the cumulative of the reduction in false negatives at state \( k \). \( \xi p_k \) and \( \xi n_k \) are given by:

\[
\xi p_k = \sum_{i=0}^{k} \varepsilon p_i, \quad \varepsilon p_i \in \{0, 1\}
\]

Eq. (53)

\[
\xi n_k = \sum_{i=0}^{k} \varepsilon n_i, \quad \varepsilon n_i \in \{0, 1\}
\]

with \( \varepsilon p_i \) and \( \varepsilon n_i \) being the events that at state \( i (i \leq k) \) the measurement errors have been reduced by one for false positive and false negatives, respectively.

Figure 42: Illustrates the progression of actual inventory and recorded inventory over time and the measurement error that constitutes the difference between them as a superposition of false positives and false negatives.

The states at which measurement errors are reduced show a binominal distribution. There are \( \Delta p_0 \) and \( \Delta n_0 \) errors that will reduce to zero over \( S \) events of demand.
Therefore, there are $S-\Delta_{p0}$ events of demand during which the measurement errors of false positives are not reduced and $S-\Delta_{n0}$ events of demand during which the measurement errors of false negatives are not reduced. Note the difference to the model with RFID where all measurement errors have already become zero when one item is remaining on the shelf (after $S-1$ events of demand).

Replenishment occurs analogous to the model with RFID when $y_{rk}$ is equal to or smaller than the threshold $r$ or when:

$$-\xi_{pk} + \xi_{nk} = r - S + (-\Delta_{p0} + \Delta_{n0}) + k$$  Eq. (54)

In the event of demand, the recorded inventory level for a model based on RFID remains either at the same level or is reduced by one. For a model based on weight-sensitive foam, the recorded inventory level remains at the same level if $\varepsilon_{pk} = 0$ and $\varepsilon_{nk} = 0$ or if $\varepsilon_{pk} = 1$ and $\varepsilon_{nk} = 1$. The recorded inventory level is reduced by one if $\varepsilon_{pk} = 1$ and $\varepsilon_{nk} = 1$, and reduced by two if $\varepsilon_{pk} = 0$ and $\varepsilon_{nk} = 1$ (Figure 42). Because the recorded inventory level may drop by two in the event of a single demand, the inventory level may be lower than $r$ when replenishment is triggered. This is the case when the recorded inventory level was $r+1$ at state $k-1$ and drops by two with the next occurrence of demand.

The descriptive model uses the same approach as the model for RFID in order to determine the expected time until replenishment. Based on the probability distributions of $\xi_{spk}$ and $\xi_{snk}$, the expected time until replenishment is calculated as the cumulative for all states $S$ and the time to go from the initial state $S$ to state $S-k$ multiplied by the probability that the recorded inventory at this state has just dropped down to or below the threshold level. In the case of RFID, the recorded inventory level at state $k-1$ is at $r+1$ and must drop down to $r$ with the occurrence of the next demand in order to trigger replenishment. In the case of weight-sensitive foam, there are two possible levels for the recorded inventory system at state $k-1$ which are $r+1$ and $r+2$. In order to trigger replenishment, the inventory level has to drop down to or below $r$ at state $k$ if it was at $r+1$ at $k-1$ or it has to drop to $r$ at state $k$ if it was at $r+1$ at $k-1$.

Let the probability that the recorded inventory level is $r+1$ at state $k-1$ be $\text{prob}(X)$. This probability is given by:
\[
\prob \left( S - \left( -\Delta p_0 + \Delta n_0 \right) - (k-1) + \left( -\xi_{p,k-1} + \xi_{n,k-1} \right) = r + 1 \mid S - (k-1) \right)
\]

\[
= \prob \left( -\xi_{p,k-1} + \xi_{n,k-1} = r - S + \left( -\Delta p_0 + \Delta n_0 \right) + k \mid S - (k-1) \right)
\]

\[
= \prob \left( \xi_{n,k-1} = r - S + \left( -\Delta p_0 + \Delta n_0 \right) + k + \xi_{p,k-1} \left( \xi_{p,k-1} = i, S - (k-1) \right) \right)
\]

\[
\cdot \prob \left( \xi_{p,k-1} = i \right)
\]

which is:

\[
\sum_i \frac{\Delta n_0}{n + i} \cdot \frac{S - \Delta n_0}{(k-1) - (n + i)} \cdot \frac{\Delta p_0}{i} \cdot \frac{S - \Delta p_0}{k - i - 1}
\]

Eq. (56)

Let the probability that the recorded inventory level is \(r+2\) at state \(k-1\) be \(\prob(\Omega)\). This probability is given by:

\[
\prob \left( S - \left( -\Delta p_0 + \Delta n_0 \right) - (k-1) + \left( -\xi_{p,k-1} + \xi_{n,k-1} \right) = r + 2 \mid S - (k-1) \right)
\]

\[
= \prob \left( -\xi_{p,k-1} + \xi_{n,k-1} = r + 1 - S + \left( -\Delta p_0 + \Delta n_0 \right) + k \mid S - (k-1) \right)
\]

\[
= \prob \left( \xi_{n,k-1} = r + 1 - S + \left( -\Delta p_0 + \Delta n_0 \right) + k + \xi_{p,k-1} \left( \xi_{p,k-1} = l, S - (k-1) \right) \right)
\]

\[
\cdot \prob \left( \xi_{p,k-1} = l \right)
\]

which is:

\[
\sum_l \frac{\Delta n_0}{n + 1 + l} \cdot \frac{S - \Delta n_0}{(k-1) - n - 1 - l} \cdot \frac{\Delta p_0}{l} \cdot \frac{S - \Delta p_0}{k - l - 1}
\]

Eq. (57)

A transition from an inventory level of \(r+1\) to an inventory level of \(r\) or below \(r\) going from \(k-1\) to \(k\) occurs if \(\epsilon_{pk} = 0\) and \(\epsilon_{nk} = 0\) or if \(\epsilon_{pk} = 1\) and \(\epsilon_{nk} = 0\) or if \(\epsilon_{pk} = 1\)
and \( \varepsilon_{nk} = 1 \). The probability of this event is the cumulative of the individual transition probabilities:

\[
\begin{align*}
prob( & \varepsilon_{p_k} = 0 \text{ and } \varepsilon_{n_k} = 0 ) \\
+ & \prob( \varepsilon_{p_k} = 1 \text{ and } \varepsilon_{n_k} = 1 ) \text{ Eq. (59)} \\
+ & \prob( \varepsilon_{p_k} = 1 \text{ and } \varepsilon_{n_k} = 0 )
\end{align*}
\]

Eq. (59) is reformulated to:

\[
\begin{align*}
prob( \varepsilon_{p_k} = 0 \mid \varepsilon_{n_k} = 0 ) \cdot prob( \varepsilon_{n_k} = 0 ) \\
+ & \prob( \varepsilon_{p_k} = 1 \mid \varepsilon_{n_k} = 1 ) \cdot prob( \varepsilon_{n_k} = 1 ) \text{ Eq. (60)} \\
+ & \prob( \varepsilon_{p_k} = 1 \mid \varepsilon_{n_k} = 0 ) \cdot prob( \varepsilon_{n_k} = 0 )
\end{align*}
\]

and under the assumption that the two variables are independent, Eq. (60) becomes:

\[
\begin{align*}
prob( \varepsilon_{p_k} = 0 ) \cdot prob( \varepsilon_{n_k} = 0 ) \\
+ & \prob( \varepsilon_{p_k} = 1 ) \cdot prob( \varepsilon_{n_k} = 1 ) \text{ Eq. (61)} \\
+ & \prob( \varepsilon_{p_k} = 1 ) \cdot prob( \varepsilon_{n_k} = 0 )
\end{align*}
\]

\( \text{Prob}(\varepsilon_{pk} = 0) \) is the probability that at state \( k \), the measurement error for false positives is reduced by one. If the measurement error for false positives, which is initially \( \Delta_{p0} \), decreases to zero in \( S \) steps, there will be \( \Delta_{p0-i} \) events over the remaining \( S-(k-1) \) steps at which the measurement error will be reduced by one. Inversely, there will be \( S-(\Delta_{p0-i})-(k-1) \) events over the remaining \( S-(k-1) \) steps during which the measurement error will not be reduced; and this corresponds to \( \text{prob}(\varepsilon_{pk} = 0) \). Hence,

\[
\prob( \varepsilon_{p_k} = 0 ) = \frac{S - (\Delta_{p0} - i) - (k - 1)}{S - (k - 1)} \text{ Eq. (62)}
\]

The probability for the event \( \varepsilon_{nk} = 0 \) is derived analogously. At state \( k-1 \), there are \( \Delta_{n0} - \xi_{nk-1} \) events remaining for the following \( S-(k-1) \) steps at which the measurement error will not be reduced. With \( \xi_{nk-1} = n+i \) and \( n = (r+1)-S+(-\Delta_{p0}+\Delta_{n0})+(k-1) \),

\[
\prob( \varepsilon_{n_k} = 0 ) = \frac{r - \Delta_{n0} + i + 1}{S - (k - 1)} \text{ Eq. (63)}
\]

The probability for the event \( \varepsilon_{pk} = 1 \) is:
\begin{equation}
prob\left(\varepsilon_{p_k} = 1\right) = \frac{\Delta_{p_0} - i}{S - (k - 1)}
\end{equation}

Eq. (64)

The probability for the event \(\varepsilon_{n_k} = 1\) is:

\begin{equation}
prob\left(\varepsilon_{n_k} = 1\right) = \frac{S - r - k + \Delta_{n_0} - i}{S - (k - 1)}
\end{equation}

Eq. (65)

where \(\Delta_{n_0}-(n+i)\) events remain at which the measurement error for false negatives is reduced by one \((n = (r+1) - \Delta_{p_0} + (\Delta_{r_0}) + (k-1))\).

Let the probability for a transition from \(r+1\) to \(r\) (Eq. (61)) be referred to as \(\text{prob}(\Phi)\).

A transition from an inventory level at \(r+2\) to \(r\) going from \(k-1\) to \(k\) occurs if \(\varepsilon_{p_k} = 1\) and \(\varepsilon_{n_k} = 0\). Let the probability of this occurrence be \(\text{prob}(\Psi)\) with:

\begin{equation}
\text{prob}(\varepsilon_{p_k} = 1) \cdot \text{prob}(\varepsilon_{n_k} = 0)
\end{equation}

Eq. (66)

which is:

\begin{equation}
\frac{\Delta_{p_0} - l}{S - (k - 1)} \cdot \frac{r - \Delta_{p_0} + 2 + l}{S - (k - 1)}
\end{equation}

Eq. (67)

Note, these expressions are only calculated for probabilities greater zero. The upper limit for \(i\) and \(l\) is \(\Delta_{p_0}\), while the lower limit depends on the state of the system. For \(k < S - 1 - \Delta_{p_0}\) the lower limit for \(i\) and \(l\) is zero. If \(k \geq S - 1 - \Delta_{p_0}\) one or more events must have already occurred at which the measurement error for false positives was reduced by one. Consequently the lower limit for \(i\) and \(l\) is given as \(\max(0, k-1-(S-\Delta_{p_0}))\).

The expected replenishment time is then:

\begin{equation}
E[t_r] = \sum_{k=1}^{S} \frac{k}{\lambda} \cdot (\text{prob}(\Phi) \cdot \text{prob}(X) + \text{prob}(\Psi) \cdot \text{prob}(\Omega))
\end{equation}

Eq. (68)

(cf. B.1 for the complete form of Eq. (68)). The expected cycle time is:

\begin{equation}
E[T_c] = E[t_r] + t_l
\end{equation}

Eq. (69)

The number of shortages depends on the demand during lead time and the actual inventory level. The actual inventory level at time \(t_r\) may be greater, smaller or equal
to \( r \) depending on the difference between the actual and the recorded inventory at time \( t_r \). Shortages occur if the demand during lead time exceeds the available inventory. The expected number of units short is given by the expected number of units short at state \( k \) multiplied by the probability that the recorded inventory has just dropped down to or below \( r \) for all \( k \). Therefore, the expected number of units short is:

\[
E[y^-] = \sum_{k=1}^{S} \left( \text{prob}(\Phi) \cdot \text{prob}(X) + \text{prob}(\Psi) \cdot \text{prob}(\Omega) \right) \cdot \sum_{x=S-k}^{\infty} \left( x - (S-k) \right) \cdot e^{-\lambda t_l} \cdot \frac{(\lambda t_l)^x}{x!} 
\]

Eq. (70)

The customer service level \( S_L \) is given by:

\[
S_L = 1 - \left( \frac{E[y^-]}{\lambda \cdot E[T_c]} \right) 
\]

Eq. (71)

The long-average cost function is formulated as:

\[
C(r, S) = \frac{T}{E[T_c]} \cdot \left( \frac{K_i \cdot S \cdot E[T_c]}{T} + \frac{c_s \cdot S \cdot E[T_c]}{T} + c_{rep} \right) 
\]

\[
+ p \cdot E[y^-] 
\]

Eq. (72)

where \( \phi \) indicates the use of mean initial measurement errors for false positives and false negatives, \( T \) the planning horizon, \( K_i \) the infrastructure cost per allocated shelf space per item per year, \( c_s \) the shelf allocation cost per item per year, \( c_{rep} \) the replenishment cost, \( p \) the penalty for shortages, \( E[y] \) the expected number of shortages per cycle, and \( y \) the inventory level. The initial measurement errors are stochastic variables. However, these variables are approximated with expected values corresponding to the results of the measurements performed in Section III.3.

### IV.4 Evaluation

In contrast to simulations, the analysis of the inventory system with mathematical models allows for a better understanding of the system’s dynamics and for simple implementation and fast computation of optimal system control parameters. However, simulations still provide a useful means to illustrate the soundness of the
computational results to the reader.

Periodic review policies are currently the most established inventory management policies at retail stores. Consequently, any inventory system intended to replace the existing inventory management policy must outperform that policy in order to be economically feasible. In the following, optimal control parameters are derived through computations for both an on-shelf inventory management system based on RFID and a system based on weight-sensitive foam. The soundness of the findings is illustrated with results generated through simulations. The performance of both systems is compared to the performance of an on-shelf inventory management policy that is based on periodic review. In addition, this thesis compares the performance of the two inventory system to provide a better understanding of how systems that introduce significant variable cost perform against systems with mostly fixed costs. Furthermore, a sensitivity analysis provides understanding on how the outputs of these models vary as the inputs change.

IV.4.1 Optimal control parameters for periodic review

IV.4.1.1 Computational results

The individual costs incurred in operating an inventory system significantly impact overall system performance and its optimal parameters. The shelf allocation cost is assumed $110 per item per year, which corresponds with the findings in II.2.5. Review cost and replenishment cost are labor costs that arise from store clerks visually inspecting and restocking shelves. The review cost is assumed $0.5 per product category per shelf. The replenishment cost is assumed $3 per product category per shelf. The penalty cost derives from the sales loss of each unit of unmet demand. The cost also accounts for lost profits on sales of other items, future sales, and special procedures used to deal with customers that are confronted with an out-of-stock situation. The penalty cost is assumed $10 per occurrence of unmet demand.

For the operation of the system, a demand of $\lambda = 10$, a review time of 1 day, and a lead time of 30min (0.02 days) are assumed. The system operates with a planning horizon of 365 days. Based on Eq. (23), the minimal costs for all sets of $(r,S,T_r)$ are computed. The set that minimizes the cost function determines the optimal values for $r$, $S$, and $T_r$ ($T_r$ is always 1 day). For the costs and parameters given above, the optimal policy for periodic review is derived with $r = 13$, $S = 17$, and $T_r = 1$. The operating cost for this policy is $3237.45$ with $0.03549$ units short per cycle. Hence, the service level is $99.63\%$ (This result along with results for other values of $c_{\text{rep}}$ and $c_r$ are given in Table 9).
**IV.4.1.2 The effect of higher demand**

An increase in demand inevitably leads to an increase in allocated shelf space and earlier triggering of replenishment. Table 7 shows how an increase in demand affects the optimal set of r and S. The service level remains high because the values for S and r are adjusted to the higher demand. Apparently, higher values for r and S impact total cost less than low values of r and S that lead to additional penalty cost.

<table>
<thead>
<tr>
<th>λ</th>
<th>C(r,S,T_r)</th>
<th>Optimal (r,S,T_r) policy</th>
<th>E[y]</th>
<th>S_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3237.45</td>
<td>(13,17,1)</td>
<td>0.03549</td>
<td>0.9963</td>
</tr>
<tr>
<td>20</td>
<td>4690.1</td>
<td>(25,29,1)</td>
<td>0.0692</td>
<td>0.99647</td>
</tr>
<tr>
<td>30</td>
<td>6066.5</td>
<td>(36,41,1)</td>
<td>0.0849665</td>
<td>0.9971</td>
</tr>
</tbody>
</table>

The maximum available shelf space allocated to a product is limited. The model assumes that an upper limit exists to the amount of shelf space allocated for a certain product. Table 8 shows optimal inventory policies where allocated shelf space is limited to 20 units. In comparison to the results in Table 7, minimal cost and units short significantly increase with λ. Note that the service level drops to about 63% because the system does not allow for the necessary adjustments of the control parameters r and S to a high demand rate.

<table>
<thead>
<tr>
<th>λ</th>
<th>C(r,S,T_r)</th>
<th>Optimal (r,S,T_r) policy</th>
<th>E[y]</th>
<th>S_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3237.45</td>
<td>(13,17,1)</td>
<td>0.03549</td>
<td>0.9963</td>
</tr>
<tr>
<td>20</td>
<td>10,593.8</td>
<td>(19,20,1)</td>
<td>1.99566</td>
<td>0.8989</td>
</tr>
<tr>
<td>30</td>
<td>41,519.3</td>
<td>(19,20,1)</td>
<td>10.6379</td>
<td>0.6383</td>
</tr>
</tbody>
</table>

**IV.4.1.3 Sensitivity analysis**

In this thesis, the mathematical models presented are defined by equations, input factors, parameters, and variables aimed at characterizing the retail replenishment process. However, input is subject to many sources of uncertainties, which may limit the meaningfulness of the findings derived from the system’s output. A sensitivity
analysis reveals the impact of a variation of an input factor on the overall system performance. The impact that system parameter variation has on the total cost function depends on the parameter and may vary. As a result, it is more important to estimate exact values for parameters with a higher effect on operating cost than those with a low effect. In order to determine the impact factor for each system input, the parameters are varied and the change in operating cost is examined.

For the system parameters given above, the overall service level remains fairly steady at a high level while the cost for operating the system increases with increasing \( c_{\text{rep}} \) and \( c_r \). The costs for review and replenishment are related because both costs represent labor costs. It is assumed that the visual inspection of shelf inventory requires less labor than moving merchandise from the backroom onto the sales floor and restocking the shelf. Consequently, \( c_{\text{rep}} > c_r \) and the ratio between the two costs remains constant because the time ratio required to perform each task is assumed constant. Table 9 reveals that with increasing \( c_{\text{rep}} \), the difference between \( S \) and \( r \) increases. According to Eq. (23), \( c_{\text{rep}} \) contributes to the total system cost through \( c_{\text{rep}} \cdot T/E[T_c] \). For increasing \( c_{\text{rep}} \), the increase in total system cost is slower if \( E[T_c] \) increases as well. However, \( E[T_c] \) increases only if \( E[n] \) increases (cf. Eq. (7)) and according to Eq. (16), \( E[n] \) increases if the difference between \( S \) and \( r \) grows. This growth in difference between \( S \) and \( r \) with increasing \( c_{\text{rep}} \) is observed in Table 9. \( S \) also contributes to the total system cost through allocated shelf space cost while \( r \) influences penalty cost that occurs through unmet demand. Because the influence of an increasing \( S \) on total system cost is significantly higher than the influence of a decreasing \( r \), only \( r \) decreases with increasing \( c_{\text{rep}} \). Review cost \( c_r \) contributes to total system cost with \( c_r \cdot E[n]/(E[n] \cdot T_r + t_i) \). Because \( E[n] \gg t_i \) for the parameters given in Table 9 and \( T_r = 1 \), the increase in total cost follows the increase of \( c_r \) in a linear manner.

System output changes linked to variations of shelf space allocation cost reveal a linear relationship. The plot of minimal cost versus \( c_s \) verifies this assumption. It also allows deriving a slope coefficient of 16, providing are rough estimate on the impact that a change of \( c_s \) has on the system’s operating cost (Figure 63 in B.2). Variations in penalty cost have a significant impact on minimal operating cost for small penalty cost \((p_c < 5)\) but the impact decreases significantly for higher penalty cost \((p_c > 30)\) (Figure 63 in B.2). The same figure also illustrates that the service level for penalty cost \( p_c > 5 \) remains at a very high level.
Table 9: Shows minimal operating cost, units short per cycle, and corresponding service level for the optimal \((r,S,T_r)\) policy for different \(c_r\) and \(c_r\) \((\lambda = 10, t_l = 0.02 \text{ days}, T_h = 365 \text{ days}, T_r = 1 \text{ day}, c_s = $110, c_p = $10)\).

<table>
<thead>
<tr>
<th>(c_r)</th>
<th>(c_r)</th>
<th>(C(r,S,T_r))</th>
<th>Optimal ((r,S,T_r)) policy</th>
<th>(E[y])</th>
<th>(S_L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.5</td>
<td>2883.2</td>
<td>(13,17)</td>
<td>0.03549</td>
<td>0.9963</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>3237.45</td>
<td>(13,17)</td>
<td>0.03549</td>
<td>0.9963</td>
</tr>
<tr>
<td>4</td>
<td>0.5</td>
<td>3588.04</td>
<td>(12,17)</td>
<td>0.04242</td>
<td>0.9955</td>
</tr>
<tr>
<td>5</td>
<td>0.5</td>
<td>3935.9</td>
<td>(12,17)</td>
<td>0.04242</td>
<td>0.9955</td>
</tr>
<tr>
<td>6</td>
<td>0.5</td>
<td>4281.55</td>
<td>(11,17)</td>
<td>0.06492</td>
<td>0.9929</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>4460.69</td>
<td>(11,17)</td>
<td>0.06492</td>
<td>0.9929</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>4818.97</td>
<td>(11,17)</td>
<td>0.06492</td>
<td>0.9929</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>5177.26</td>
<td>(11,17)</td>
<td>0.06492</td>
<td>0.9929</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>5535.54</td>
<td>(11,17)</td>
<td>0.06492</td>
<td>0.9929</td>
</tr>
<tr>
<td>9</td>
<td>1.5</td>
<td>5257.23</td>
<td>(12,26)</td>
<td>0.09177</td>
<td>0.9821</td>
</tr>
</tbody>
</table>

**IV.4.2 Optimal control policies for RFID systems with imperfect state information**

**IV.4.2.1 Computational results**

In order to permit for comparable results among different kinds of inventory management systems, each system assumes the same initial costs for shelf space allocation ($110 per item per year), shelf replenishment ($3 per replenishment), and penalty cost for stock-outs ($10 per unit short). In contrast to periodic review, no labor costs arise for the visual inspection of shelf stocks. Instead, the system accounts for RFID infrastructure cost such as RFID antennas, reader and data processing with an infrastructure cost \(K_I\) of $12 per allocated shelf space per year. The RFID tag cost represent a variable cost and contribute to the review cost with $0.2 per item sold. The initial difference \(\Delta_0\) between actual inventory and recorded inventory caused through shielding is represented by read rate \(\varphi\). Consequently, \(\Delta_0 = S \cdot (1 - \varphi)\). According to III.2.3, read rates for high frequency RFID systems are around 90%-100%. Nevertheless, for the analysis and understanding of the system dynamics, read rates ranging from 10%-100% are taken into consideration. Figure 43 shows simulation results on the progression of actual inventory and recorded inventory over time with a read rate of 82%. The progression of the error follows the relationship given in Eq. (31). Note how the error progresses over time and becomes zero before the actual inventory level reaches \(r = 1\) after \(S\)-1 steps.
Figure 43: Illustrates the progression of actual inventory (dashed) and recorded inventory (solid) over time. Once the recorded inventory reaches $r = 1$, replenishment is triggered and after a lead time, the actual inventory is raised to 11 and the recorded inventory to 9 to start a new cycle.

The inventory management system is operated under the same conditions as the periodic inventory system. Demand rate $\lambda$ is 10, the lead time is 30 min (0.02 days) and the planning horizon is 365 days. The minimal costs for all sets of $(r, S)$ are computed according to Eq. (48). The set that minimizes the cost function represents the optimal values for $r$ and $S$. Optimal policies for RFID inventory management systems at different read rates are given in Table 10 for tag cost of $0.2$ and in Table 11 for tag cost of $0.2$. 
Table 10: Shows the minimal operating cost, units short per cycle, 
corresponding service level, and additional replenishments due to early 
triggering of the system for the optimal (r,S) policy at read rate $\varphi$, $c_{tag} = 0.2$, 
and $\lambda = 10$ ( $t_l = 0.02$ days, $T_h = 365$ days, $K_i = 12$, $c_{rep} = 3$, $c_s = 110$, $c_p = 10$).

<table>
<thead>
<tr>
<th>$\varphi$</th>
<th>$C_\varphi(r,S)$</th>
<th>Optimal (r,S) policy</th>
<th>$E[y]$</th>
<th>$S_L$</th>
<th># of add rep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3212.56</td>
<td>(1,11)</td>
<td>0.01873</td>
<td>0.99809</td>
<td>0</td>
</tr>
<tr>
<td>0.9</td>
<td>3217.52</td>
<td>(1,11)</td>
<td>0.01698</td>
<td>0.99829</td>
<td>3.5400</td>
</tr>
<tr>
<td>0.8</td>
<td>3225.04</td>
<td>(1,11)</td>
<td>0.01520</td>
<td>0.99848</td>
<td>7.9698</td>
</tr>
<tr>
<td>0.7</td>
<td>3236.27</td>
<td>(1,11)</td>
<td>0.01339</td>
<td>0.99868</td>
<td>13.6581</td>
</tr>
<tr>
<td>0.6</td>
<td>3253.09</td>
<td>(1,11)</td>
<td>0.01157</td>
<td>0.99889</td>
<td>21.2370</td>
</tr>
<tr>
<td>0.5</td>
<td>3278.87</td>
<td>(1,11)</td>
<td>0.00971</td>
<td>0.99909</td>
<td>31.8366</td>
</tr>
<tr>
<td>0.4</td>
<td>3337.90</td>
<td>(1,12)</td>
<td>0.00713</td>
<td>0.99930</td>
<td>46.5561</td>
</tr>
<tr>
<td>0.3</td>
<td>3405.57</td>
<td>(1,12)</td>
<td>0.00538</td>
<td>0.99951</td>
<td>70.8463</td>
</tr>
<tr>
<td>0.2</td>
<td>3564.60</td>
<td>(1,13)</td>
<td>0.003313</td>
<td>0.99970</td>
<td>112.4740</td>
</tr>
<tr>
<td>0.1</td>
<td>3677.93</td>
<td>(0,12)</td>
<td>0.20000</td>
<td>0.97560</td>
<td>0</td>
</tr>
</tbody>
</table>

In comparison to optimal policies for periodic review, optimal policies for RFID 
inventory management systems show low replenishment threshold values r. 
Continuous monitoring combined with a short lead time for replenishments and 
moderate demand rates allow for a low replenishment threshold because inventory

Table 11: Shows the minimal operating cost, units short per cycle, 
corresponding service level, and additional replenishments due to early 
triggering of the system for the optimal (r,S) policy at read rate $\varphi$, $c_{tag} = 0.1$, 
and $\lambda = 10$ ( $t_l = 0.02$ days, $T_h = 365$ days, $K_i = 12$, $c_{rep} = 3$, $c_s = 110$, $c_p = 10$).

<table>
<thead>
<tr>
<th>$\varphi$</th>
<th>$C_\varphi(r,S)$</th>
<th>Optimal (r,S) policy</th>
<th>$E[y]$</th>
<th>$S_L$</th>
<th># of add rep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2847.56</td>
<td>(1,11)</td>
<td>0.01873</td>
<td>0.99809</td>
<td>0</td>
</tr>
<tr>
<td>0.9</td>
<td>2852.52</td>
<td>(1,11)</td>
<td>0.01698</td>
<td>0.99829</td>
<td>3.5400</td>
</tr>
<tr>
<td>0.8</td>
<td>2860.04</td>
<td>(1,11)</td>
<td>0.01520</td>
<td>0.99848</td>
<td>7.97</td>
</tr>
<tr>
<td>0.7</td>
<td>2871.27</td>
<td>(1,11)</td>
<td>0.01339</td>
<td>0.99868</td>
<td>13.66</td>
</tr>
<tr>
<td>0.6</td>
<td>2888.09</td>
<td>(1,11)</td>
<td>0.01157</td>
<td>0.99889</td>
<td>21.24</td>
</tr>
<tr>
<td>0.5</td>
<td>2913.87</td>
<td>(1,11)</td>
<td>0.00971</td>
<td>0.99909</td>
<td>31.84</td>
</tr>
<tr>
<td>0.4</td>
<td>2972.9</td>
<td>(1,12)</td>
<td>0.00713</td>
<td>0.99930</td>
<td>46.56</td>
</tr>
<tr>
<td>0.3</td>
<td>3040.57</td>
<td>(1,12)</td>
<td>0.00538</td>
<td>0.99951</td>
<td>70.85</td>
</tr>
<tr>
<td>0.2</td>
<td>3199.60</td>
<td>(1,13)</td>
<td>0.00331</td>
<td>0.99971</td>
<td>112.47</td>
</tr>
<tr>
<td>0.1</td>
<td>3318.92</td>
<td>(0,12)</td>
<td>0.20000</td>
<td>0.97560</td>
<td>0</td>
</tr>
</tbody>
</table>

In comparison to optimal policies for periodic review, optimal policies for RFID 
inventory management systems show low replenishment threshold values r. 
Continuous monitoring combined with a short lead time for replenishments and 
moderate demand rates allow for a low replenishment threshold because inventory
levels that reach the threshold are detected instantly. Due to timely replenishments, the probability for stock-outs remains low despite a low threshold level. A low threshold reduces the total number of replenishments and the costs incurred thereof.

Additionally, Table 10 and Table 11 demonstrate that the number of units short, the corresponding service level, and the total number of replenishments increase as the read rate decreases. If the recorded inventory level reaches the threshold while the actual inventory still shows an inventory level higher than the threshold, an early triggering of replenishment occurs. Early triggering of replenishment is due to the difference in actual inventory and recorded inventory. With an increasing initial difference, the probability that the recorded inventory reaches the threshold and triggers replenishment before all shielding errors have been eliminated from the system increases as well. Consequently, the higher the actual inventory level when replenishment is triggered, the lower the probability for stock-outs. In addition, the earlier the recorded inventory reaches the threshold (in contrast to when the actual inventory level would have reached the threshold) the higher the number of additional replenishments. The number of additional replenishments is computed according to Eq. (51). To some extent, the system tries to reduce the incurrence of costs due to additional replenishment by increasing the allocated shelf space to carry higher inventory stocks.

For a read rate of only 10%, the behavior of the system differs from the pattern described above. With such a low initial recorded inventory level, it becomes more economical to reduce the threshold to zero and account for penalty costs due to units short rather than for additional replenishments. Because elimination of measurement errors from the system occurs with the removal of the last item, the recorded and actual inventory show the same inventory level right before shelf depletion. Consequently, an early triggering of the replenishment process cannot occur. However, with a replenishment threshold of zero, the system cannot meet any demand that occurs during the replenishment lead time. Therefore, the number of units short is considerably higher than for other read rates.

The accuracy of the computational results is easily verified with simulations. A MATLAB simulation with preset values for r and S simulates the reduction in actual and recorded inventory. The random inter-arrival times of customers are generated according to a Poisson process with $\lambda = 10$. With each arriving customer, the actual inventory is reduced by one. The progression of the recorded inventory is a function of arriving customers and residual measurement errors for the steps remaining until depletion of the shelf. Specifically, a random number is generated each time demand occurs. If that random number is larger than the residual measurement errors divided
by the number of steps remaining, the recorded inventory level is reduced by one. Otherwise, it stays at the same level and the number of measurement errors is reduced. Once the recorded inventory level reaches the threshold, demand continues to occur until the cumulative of the inter-arrival times exceeds the lead time. After replenishment, the actual inventory level is reset to $S$ and the recorded inventory level to $S \cdot \varphi$ ($\varphi$ being the read rate) and a new cycle begins. Over a certain number of consecutive days, the simulation counts the number of units short and the times until replenishments occur. These values are divided by the total number of replenishment cycles at the end of the simulation run to derive the mean number of units short per cycle and the mean time until replenishment is triggered.

Each simulation runs for 10,000 consecutive days, which corresponds to 22.4 years of operating the inventory system. Ten simulations runs are conducted for each set of control parameters to derive the mean value and standard deviation for the time until replenishment as well as for units short. The overall similarities between the calculated and the simulated results in Table 12 indicate that the mathematical models are correct. Although the confidence intervals are not as small as desired, more extensive simulations have revealed that even with multiple simulation runs and simulation periods of 25,000 consecutive days, the results do not become more meaningful (cf. Appendix B.2). A simulation period of 10,000 consecutive days already requires several minutes to complete for one set of control parameters. There are $n \cdot (n-1)$ control sets where $n$ forms the upper limit for $S$. Because the optimal value for $S$ is a priori unknown, $n$ needs to be sufficiently high. For each $S$, there is an optimal value $r$ between 0 and $(S-1)$ that minimizes the cost function. In order to find the optimal set for $r$ and $S$ that minimizes the cost function, $n \cdot (n-1)$ control sets need to be evaluated. Note that when compared to the cumbersome simulations, calculations for individual sets of control parameters are completed within seconds.
Table 12: Shows the calculated values and simulation results for expected time until replenishment and expected number of units short for different (r,S) policies and read rates. Ten simulations are carried out and each runs for 10,000 consecutive days.

<table>
<thead>
<tr>
<th>φ</th>
<th>(r,S)</th>
<th>E[t_r] (Calc)</th>
<th>μ of t_r (Sim)</th>
<th>σ of t_r (Sim)</th>
<th>E[y-] (Calc)</th>
<th>μ of y- (Sim)</th>
<th>σ of y- (Sim)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>(1,10)</td>
<td>0.90000</td>
<td>0.90007</td>
<td>0.00288</td>
<td>0.01873</td>
<td>0.01880</td>
<td>0.00103</td>
</tr>
<tr>
<td>1.0</td>
<td>(2,10)</td>
<td>0.80000</td>
<td>0.80028</td>
<td>0.00246</td>
<td>0.00121</td>
<td>0.00109</td>
<td>0.00034</td>
</tr>
<tr>
<td>1.0</td>
<td>(1,11)</td>
<td>1.00000</td>
<td>0.99851</td>
<td>0.00306</td>
<td>0.01873</td>
<td>0.01834</td>
<td>0.00132</td>
</tr>
<tr>
<td>1.0</td>
<td>(2,11)</td>
<td>0.90000</td>
<td>0.89987</td>
<td>0.00275</td>
<td>0.00121</td>
<td>0.00122</td>
<td>0.00037</td>
</tr>
<tr>
<td>0.9</td>
<td>(1,10)</td>
<td>0.88889</td>
<td>0.88742</td>
<td>0.00267</td>
<td>0.01678</td>
<td>0.01711</td>
<td>0.00098</td>
</tr>
<tr>
<td>0.9</td>
<td>(2,10)</td>
<td>0.77778</td>
<td>0.77718</td>
<td>0.00278</td>
<td>0.00095</td>
<td>0.00100</td>
<td>0.00023</td>
</tr>
<tr>
<td>0.9</td>
<td>(1,11)</td>
<td>0.99000</td>
<td>0.99051</td>
<td>0.00370</td>
<td>0.01698</td>
<td>0.01765</td>
<td>0.00151</td>
</tr>
<tr>
<td>0.9</td>
<td>(2,11)</td>
<td>0.88000</td>
<td>0.87977</td>
<td>0.00306</td>
<td>0.00098</td>
<td>0.00082</td>
<td>0.00015</td>
</tr>
<tr>
<td>0.8</td>
<td>(1,10)</td>
<td>0.87500</td>
<td>0.88629</td>
<td>0.03487</td>
<td>0.01481</td>
<td>0.01491</td>
<td>0.00129</td>
</tr>
<tr>
<td>0.8</td>
<td>(2,10)</td>
<td>0.75000</td>
<td>0.74816</td>
<td>0.00209</td>
<td>0.00072</td>
<td>0.00067</td>
<td>0.00015</td>
</tr>
<tr>
<td>0.8</td>
<td>(1,11)</td>
<td>0.97778</td>
<td>0.97822</td>
<td>0.00335</td>
<td>0.01520</td>
<td>0.01466</td>
<td>0.00132</td>
</tr>
<tr>
<td>0.8</td>
<td>(2,11)</td>
<td>0.85556</td>
<td>0.85622</td>
<td>0.00289</td>
<td>0.00077</td>
<td>0.00089</td>
<td>0.00266</td>
</tr>
</tbody>
</table>

While the minimal operating cost for a RFID inventory system depends on read rate performance, the operating cost for periodic review systems do not. Therefore, the minimal costs for periodic review form horizontal lines in Figure 44. The line for minimal operating cost for a periodic review system moves vertical if c_r varies (cf. Eq. (17)). According to Table 10, the impact of the read rate on the operating costs for RFID systems is significant. Therefore, the read rate has to be taken into account for the evaluation and comparison of RFID and periodic review inventory systems. Figure 44 illustrates the change in minimal operating cost for RFID systems as the read rate changes. Two plots are given for RFID systems operating with tag cost of $0.2 and $0.1, respectively. Minimal costs for each read rate and tag cost are calculated and plotted. The difference between the two lines representing the two RFID inventory systems derives from the difference in tag cost. The tag cost contributes with c_tag · min(S, λ·E[T_c]) to the system cost. A change of Δc_tag affects the entire cost function because the control parameters are affected through min(S, λ·E[T_c]). The impact on the shape is minimal in comparison to the amount of vertical displacement. More significant changes in shape are observed for higher demand rates (cf. Figure 62 in B.2).

Figure 44 reveals that for moderate read rates and tag cost of $0.1, the RFID inventory system operates at significantly lower cost than the periodic review system.
with review cost of $0.5. For a RFID system with tag cost of $0.2, the system outperforms periodic review by a small margin only for high read rates. For low read rates, the RFID system performs worse. However, small changes to the system parameters may have a significant impact on the outcome of the performance analysis. Figure 44 also shows how the minimal operating costs for the RFID systems compare to periodic review system with review cost of $0.1 and $0.25, respectively.

![Graph showing minimal costs for periodic review and RFID systems with respect to read rate and tag cost of $0.1 or $0.2 and review cost of $1, $0.5, $0.25 ($\lambda = 10$).](image)

**Figure 44**: Shows minimal costs for periodic review and RFID systems with respect to read rate and tag cost of $0.1 or $0.2 and review cost of $1, $0.5, $0.25 ($\lambda = 10$).

**IV.4.2.2 The effect of higher demand**

Similar to periodic review inventory systems, an increase in demand leads to an increase in allocated shelf space for RFID inventory systems and inevitably to an increase in minimal operating cost. However, if an upper limit for allocated shelf space of 20 units is assumed, Table 13 and Table 14 show that the only policies affected are the ones that show lowest read rate performances. Additionally, even though threshold $r$ remains rather low, the service level remains high despite higher demand. This is due to the continuous monitoring strategy and timely replenishment. In fact, the threshold drops with decreasing read rates because lower read rates increase the likelihood of an early replenishment. Hence, the requirement for safety stocks is minimal.
Table 13: Shows the minimal operating cost, units short per cycle, corresponding service level, and additional replenishments for the optimal (r,S) policy at read rate $\phi$ and $\lambda = 20$ ($c_{tag} = \$0.2$).

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>$C_\phi(r,S)$</th>
<th>Optimal (r,S) policy</th>
<th>$E[y]$</th>
<th>$S_L$</th>
<th># of add rep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4972.09</td>
<td>(2,15)</td>
<td>0.00877</td>
<td>0.99971</td>
<td>0</td>
</tr>
<tr>
<td>0.9</td>
<td>4983.99</td>
<td>(2,15)</td>
<td>0.00764</td>
<td>0.99975</td>
<td>5.87</td>
</tr>
<tr>
<td>0.8</td>
<td>5013.36</td>
<td>(2,16)</td>
<td>0.00576</td>
<td>0.99979</td>
<td>16.79</td>
</tr>
<tr>
<td>0.7</td>
<td>5044.75</td>
<td>(2,15)</td>
<td>0.00462</td>
<td>0.99985</td>
<td>31.26</td>
</tr>
<tr>
<td>0.6</td>
<td>5089.26</td>
<td>(2,16)</td>
<td>0.00329</td>
<td>0.99989</td>
<td>46.09</td>
</tr>
<tr>
<td>0.5</td>
<td>5111.68</td>
<td>(1,16)</td>
<td>0.03528</td>
<td>0.99870</td>
<td>32.92</td>
</tr>
<tr>
<td>0.4</td>
<td>5136.75</td>
<td>(1,17)</td>
<td>0.02870</td>
<td>0.99893</td>
<td>42.47</td>
</tr>
<tr>
<td>0.3</td>
<td>5185.75</td>
<td>(1,16)</td>
<td>0.02073</td>
<td>0.99932</td>
<td>79.00</td>
</tr>
<tr>
<td>0.2</td>
<td>5304.46</td>
<td>(1,18)</td>
<td>0.01388</td>
<td>0.99952</td>
<td>105.64</td>
</tr>
<tr>
<td>0.1</td>
<td>5821.55</td>
<td>(1,25)</td>
<td>0.00663</td>
<td>0.99972</td>
<td>128.55</td>
</tr>
</tbody>
</table>

Table 14: Shows the minimal operating cost, units short per cycle, corresponding service level, and additional replenishments for the optimal (r,S) policy at read rate $\phi$ and $\lambda = 30$ ($c_{tag} = \$0.2$).

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>$C_\phi(r,S)$</th>
<th>Optimal (r,S) policy</th>
<th>$E[y]$</th>
<th>$S_L$</th>
<th># of add rep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6511.95</td>
<td>(3,19)</td>
<td>0.00379</td>
<td>0.99993</td>
<td>0</td>
</tr>
<tr>
<td>0.9</td>
<td>6537.61</td>
<td>(2,19)</td>
<td>0.02186</td>
<td>0.99958</td>
<td>8.43034</td>
</tr>
<tr>
<td>0.8</td>
<td>6542.99</td>
<td>(2,19)</td>
<td>0.01717</td>
<td>0.99967</td>
<td>19.4425</td>
</tr>
<tr>
<td>0.7</td>
<td>6560.33</td>
<td>(2,18)</td>
<td>0.01435</td>
<td>0.99975</td>
<td>32.0524</td>
</tr>
<tr>
<td>0.6</td>
<td>6595.26</td>
<td>(2,20)</td>
<td>0.00989</td>
<td>0.99981</td>
<td>45.4602</td>
</tr>
<tr>
<td>0.5</td>
<td>6638.64</td>
<td>(2,19)</td>
<td>0.00744</td>
<td>0.99987</td>
<td>70.8789</td>
</tr>
<tr>
<td>0.4</td>
<td>6753.3</td>
<td>(2,19)</td>
<td>0.00449</td>
<td>0.99993</td>
<td>115.215</td>
</tr>
<tr>
<td>0.3</td>
<td>6819.9</td>
<td>(1,20)</td>
<td>0.04532</td>
<td>0.99913</td>
<td>75.4964</td>
</tr>
<tr>
<td>0.2</td>
<td>6908.25</td>
<td>(1,23)</td>
<td>0.03180</td>
<td>0.99933</td>
<td>91.8025</td>
</tr>
<tr>
<td>0.1</td>
<td>7236.81</td>
<td>(1,25)</td>
<td>0.01487</td>
<td>0.99972</td>
<td>189.048</td>
</tr>
</tbody>
</table>
Table 15: Shows the minimal operating cost, units short per cycle, corresponding service level, and additional replenishments for the optimal (r,S) policy at read rate $\phi$ and $\lambda = 20$ ($c_{\text{tag}} = $0.1).

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>$C_{\phi}(r,S)$</th>
<th>Optimal (r,S) policy</th>
<th>$E[y]$</th>
<th>$S_L$</th>
<th># of add rep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4242.09</td>
<td>(2,15)</td>
<td>0.00877</td>
<td>0.99971</td>
<td>0</td>
</tr>
<tr>
<td>0.9</td>
<td>4253.99</td>
<td>(2,15)</td>
<td>0.00764</td>
<td>0.99975</td>
<td>5.87</td>
</tr>
<tr>
<td>0.8</td>
<td>4283.36</td>
<td>(2,16)</td>
<td>0.00576</td>
<td>0.99979</td>
<td>16.79</td>
</tr>
<tr>
<td>0.7</td>
<td>4314.75</td>
<td>(2,15)</td>
<td>0.00462</td>
<td>0.99985</td>
<td>31.26</td>
</tr>
<tr>
<td>0.6</td>
<td>4359.26</td>
<td>(2,16)</td>
<td>0.00329</td>
<td>0.99989</td>
<td>46.09</td>
</tr>
<tr>
<td>0.5</td>
<td>4381.68</td>
<td>(1,16)</td>
<td>0.03528</td>
<td>0.99873</td>
<td>32.92</td>
</tr>
<tr>
<td>0.4</td>
<td>4406.75</td>
<td>(1,17)</td>
<td>0.02870</td>
<td>0.99893</td>
<td>42.47</td>
</tr>
<tr>
<td>0.3</td>
<td>4455.75</td>
<td>(1,16)</td>
<td>0.02073</td>
<td>0.99932</td>
<td>79.00</td>
</tr>
<tr>
<td>0.2</td>
<td>4574.46</td>
<td>(1,18)</td>
<td>0.01388</td>
<td>0.99952</td>
<td>105.64</td>
</tr>
<tr>
<td>0.1</td>
<td>5292.33</td>
<td>(1,18)</td>
<td>0.004706</td>
<td>0.99989</td>
<td>357.06</td>
</tr>
</tbody>
</table>

Table 16: Shows the minimal operating cost, units short per cycle, corresponding service level, and additional replenishments for the optimal (r,S) policy at read rate $\phi$ and $\lambda = 30$ ($c_{\text{tag}} = $0.1).

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>$C_{\phi}(r,S)$</th>
<th>Optimal (r,S) policy</th>
<th>$E[y]$</th>
<th>$S_L$</th>
<th># of add rep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5416.95</td>
<td>(3,19)</td>
<td>0.00379</td>
<td>0.99993</td>
<td>0</td>
</tr>
<tr>
<td>0.9</td>
<td>5442.61</td>
<td>(2,19)</td>
<td>0.02186</td>
<td>0.99958</td>
<td>8.43</td>
</tr>
<tr>
<td>0.8</td>
<td>5447.99</td>
<td>(2,19)</td>
<td>0.01717</td>
<td>0.99967</td>
<td>19.44</td>
</tr>
<tr>
<td>0.7</td>
<td>5465.33</td>
<td>(2,18)</td>
<td>0.01435</td>
<td>0.99975</td>
<td>32.05</td>
</tr>
<tr>
<td>0.6</td>
<td>5500.26</td>
<td>(2,20)</td>
<td>0.00989</td>
<td>0.99981</td>
<td>45.46</td>
</tr>
<tr>
<td>0.5</td>
<td>5543.64</td>
<td>(2,19)</td>
<td>0.00744</td>
<td>0.99987</td>
<td>70.88</td>
</tr>
<tr>
<td>0.4</td>
<td>5658.3</td>
<td>(2,19)</td>
<td>0.00449</td>
<td>0.99993</td>
<td>115.22</td>
</tr>
<tr>
<td>0.3</td>
<td>5724.9</td>
<td>(1,20)</td>
<td>0.04531</td>
<td>0.99913</td>
<td>75.49</td>
</tr>
<tr>
<td>0.2</td>
<td>5835.87</td>
<td>(1,20)</td>
<td>0.02779</td>
<td>0.99952</td>
<td>143.25</td>
</tr>
<tr>
<td>0.1</td>
<td>6731.92</td>
<td>(1,20)</td>
<td>0.00947</td>
<td>0.99989</td>
<td>474.35</td>
</tr>
</tbody>
</table>

For higher demand rates such as $\lambda = 20$ and $\lambda = 30$ RFID systems show lower operating cost than periodic review for tag cost of $0.1$, while the operating cost for tag cost of $0.2$ is higher. Figure 45 illustrates how the difference in operating costs between the RFID and the periodic review system increases with higher demand rates. However, for low demand and low review costs periodic review inventory
management systems perform well and may not justify replacing the existing and established inventory management policy.

Figure 45: Shows the minimal operating costs for periodic review systems and RFID systems at different demand rates. For demand rates equal to and higher than $\lambda = 20$, RFID systems with $c_{\text{tag}} = $0.2 perform worse than periodic review, while RFID system with $c_{\text{tag}} = $0.1 perform better.

Adjusting the control parameter for higher demand allows maintaining high customer service levels. For RFID systems, the control parameters are lower than for periodic review system. This characteristic becomes important when the maximum amount of allocated shelf space is limited and restricts the necessary adjustments of $S$ to meet higher demand (Table 8 shows how the service level drops for periodic review systems that are confronted with high demand but have restricted shelf space of 20 items. Table 13 shows that the service level for RFID systems with read rates higher than 10% remains unaffected by this restriction). Consequently, for high demand and limited shelf space, RFID inventory management systems with low tag cost are operated at significantly lower cost than periodic review systems while maintaining a high service level.

**IV.4.2.3 Sensitivity analysis**

In order to identify the critical input factors that lead to a significant change of the system’s output if varied slightly, the effects on the system output are examined for variations of all inputs. The estimations for most input factors are subject to
uncertainty. Therefore, it is important to understand which input factors need to be estimated more carefully so that the mathematical model produces accurate results. For variations of the input parameters shelf allocation cost $c_s$, shelf replenishment cost $c_{rep}$, and infrastructure cost $K_I$ the change in operating cost is approximately linear. Plotting the variation of each individual parameter and the corresponding change in minimal cost, a slope coefficient for each parameter is derived. These slope coefficients provide a rough indication on the system’s sensitivity to a change of a specific parameter. For the plot (Figure 64 in B.2):

- Minimal cost versus $c_{rep}$, a slope coefficient of approximately 350 is derived for variations of $c_{rep}$ between 0 and 9.
- Minimal cost versus $\lambda$, a slope coefficient of approximately 170 is derived for variations of $\lambda$ between 10 and 30.
- Minimal cost versus $c_s$, a slope coefficient of approximately 9 is derived for variations of $c_s$ between 50 and 350.
- Minimal cost versus $K_I$, a slope coefficient of approximately 9 is derived for variations of $K_I$ between 0 and 150.

The higher the slope coefficient, the more severe is the influence of a system parameter on minimal process cost. Consequently, during the data gathering process prior to system evaluation more time and effort should be assigned to a detailed analysis of those parameters that show high slope coefficients.

Penalty cost differs from other costs because it shows a highly non-linear relation to operating cost. Penalty cost consists of several different components. These components include: lost sales for each unit of unmet demand, lost profits on sales of other items, loss on future sales, and costs arising from dealing with customers that face stock-outs. Estimating these components exactly may be challenging. However, if the penalty cost is assumed high, variations to this parameter do not result in a significant change of the output (Figure 46). In contrast to operating cost, an increase in penalty cost does affect the number of units short and the customer service level, respectively. Figure 47 illustrates how the number of units short drops with increasing penalty cost.
Figure 46: Shows the increase in minimal operating cost as the penalty cost increases (given for different read rates).

Figure 47: Shows the reduction in units short with increasing penalty cost (given for different read rates).
The sensitivity analysis illustrates that some system parameters need to be evaluated carefully in order to obtain conclusive results from an inventory system performance analysis. A profound understanding of the system’s response to changes of the inputs becomes even more important if different systems are compared.

**IV.4.3 Optimal control parameters for weigh-sensitive foam systems with imperfect state information**

**IV.4.3.1 Computational results**

The inventory management system based on weight-sensitive foam operates under the same constraints as the periodic review and RFID inventory system. Shelf space allocation cost, shelf replenishment cost, and penalty cost for stock-outs remain the same. In contrast to RFID systems that account for both, fix infrastructure cost and variable cost for tags, foam-based systems do not have any variable cost components. However, foam-based systems account for fix infrastructure costs that include costs for foam, a data processing unit, and communication technology.

Because foam becomes a part of the shelf after installation, the infrastructure cost \( K_I \) is treated similar to shelf allocation cost as a cost per item per year. The infrastructure cost \( K_I \) is conservatively estimated at $7.5 per item per year. (This estimate is based on the considerations for manufacturing costs in III.3.8, a significantly higher sales price (approximately 500%), maintenance cost, software cost, and an estimate on how many items may be displayed on one running meter of foam. This estimate is subject to a high degree of uncertainty. Therefore, the infrastructure cost for foam-based systems is treated as a variable when compared to RFID-based systems).

The findings in III.3 suggest that measurement errors for foam-based systems are in the range of 0% to 10% for both false positives and false negatives. Additionally, false positives tend to be much smaller than false negatives. Nevertheless, in order to provide a better understanding of the system dynamics, larger detection errors are examined as well. Similar to RFID, the number of measurement errors depends on the number of items on display. False negatives occur when an item covers a sensor but the sensor does not report a load; and false positives are primarily caused by poor alignment of products on display. Poor alignment of products partially covers neighboring sensors. Both measurement errors cause an initial difference between the recorded inventory and the actual inventory \( y_t = y_a - \Delta_n + \Delta_p \). The initial error for false negatives, \( \Delta_n \), is given by \( S \cdot d_n \) (\( d_n \) being the detection rate for false negatives); and the initial error for false positives, \( \Delta_p \), is given by \( S \cdot d_p \) (\( d_p \) being
the detection rate for false positives). The recorded inventory level progresses according to the relationship described in Eq. (52). Detection errors will disappear with the removal of the last item after $S$ events of demand and under the assumption that no replenishment occurred. Each individual item may cause either one false positive or one false negative, or both errors at the same time, or none at all. Consequently, the removal of an item may keep the recorded inventory level at the same level (for a false negative), reduce it by one (for no errors or for a concurrent false negative and false positive) or reduce it by two (for a false positive). Figure 48 shows simulation results on the progression of actual inventory and recorded inventory over time with an initial measurement error for false negatives of 3 ($d_n$ equal to 0.33%) and an initial measurement error for false positives of 1 ($d_p$ equal to 0.1%).

Figure 48: Illustrates the progression of actual inventory (dashed) and recorded inventory (solid) over time. The initial actual inventory level is 10. The initial recorded inventory level is 8 with initial measurement errors of 3 for false negatives and 1 for false positives.

In order to derive comparable results, the foam-based inventory management system is operated under the same constraints as the periodic review and the RFID system. The demand rate $\lambda$ is 10, the lead time $t_r$ is 30min, and the planning horizon is 365 days. The minimal costs for all sets of $(r,S)$ are computed according to Eq. (72). Optimal policies for foam-based systems and different detection rates for false
negatives and false positive are given in Table 17.

Table 17: Shows optimal policies for different detection rates of false negatives and positives, minimal operating cost, units short per cycle, and corresponding service level ($\lambda = 10$, $t_l = 0.02$ days, $T_h = 365$ days, $K_1 = $7.5, $c_{rep} = $3, $c_s = $110, $c_p = $10).

<table>
<thead>
<tr>
<th>$d_n$</th>
<th>$d_p$</th>
<th>$C_v(r,S)$</th>
<th>Optimal (r,S) policy</th>
<th>$E[y']$</th>
<th>$S_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>2433.06</td>
<td>(1,11)</td>
<td>0.01873</td>
<td>0.99809</td>
</tr>
<tr>
<td>0.0</td>
<td>0.1</td>
<td>2476.74</td>
<td>(2,11)</td>
<td>0.00439</td>
<td>0.99959</td>
</tr>
<tr>
<td>0.0</td>
<td>0.2</td>
<td>2467.41</td>
<td>(2,11)</td>
<td>0.00726</td>
<td>0.99931</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0</td>
<td>2442.15</td>
<td>(1,11)</td>
<td>0.01555</td>
<td>0.99844</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>2485.07</td>
<td>(1,11)</td>
<td>0.03111</td>
<td>0.99685</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
<td>2487.65</td>
<td>(2,11)</td>
<td>0.00606</td>
<td>0.99943</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0</td>
<td>2456.83</td>
<td>(1,11)</td>
<td>0.01262</td>
<td>0.99877</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>2492.77</td>
<td>(1,11)</td>
<td>0.02716</td>
<td>0.99729</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>2509.65</td>
<td>(2,12)</td>
<td>0.00478</td>
<td>0.99952</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0</td>
<td>2479.21</td>
<td>(1,11)</td>
<td>0.00995</td>
<td>0.99905</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>2505.91</td>
<td>(1,11)</td>
<td>0.02337</td>
<td>0.99773</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
<td>2533.23</td>
<td>(1,11)</td>
<td>0.03667</td>
<td>0.99637</td>
</tr>
</tbody>
</table>

The findings for minimal operating costs from Table 17 and their correlation with detection rates for false negatives and false positives are illustrated in Figure 49. Note that the minimal operating costs are significantly lower than comparable cost for periodic review systems ($3237.45) and RFID systems ($3212.56 with read rate of 100% and tag cost of $0.1).
Figure 49: Shows minimal costs as functions of detection rates for false negatives and false positives ($d_p$).

Similar to optimal policies for RFID systems, optimal control parameters for foam-based systems show low replenishment threshold values $r$. While false negatives increase the likelihood for early triggering of replenishment, false positives have the opposite effect. Therefore, for a high detection rate of false negatives and a low rate of false positives, the threshold value $r$ is at one. This threshold value rises to two for a low detection rate of false negatives and a high rate of false positives. Because a high detection rate of false positives results in late triggering of replenishments (if these errors are not partly compensated by false negatives), the probability for stockouts increases. In order to balance additional costs arising from penalty costs due to unmet demand, the threshold is increased. Increasing the threshold does reduce the amount of penalty cost induced, but also results in more frequent replenishments that increase the overall cost for replenishment. As a result of raising the threshold for higher numbers of false positives, the overall service level remains at a fairly high level.

Foam-based systems that are not subject to false positives show very similar results for optimal policies and expected number of units short as RFID systems for read rates that correspond to the detection rate of false negatives (cf. Table 10 and Table 17). Expected numbers for units short are slightly lower for foam-based systems than for RFID systems. This is explained by the fact that errors in the recorded
The inventory of a foam-based system may remain in the system until \( S \) occurrences of demand while for RFID systems, the recorded inventory becomes identical with the actual inventory the latest after \( S-1 \) occurrences of demand. Consequently, the probability that the system is still subject to errors when the recorded inventory level reaches the threshold is larger for foam-based systems than for RFID by a small fraction. Therefore, the likelihood for early triggering of replenishment is higher for foam-based systems. Early triggering results in a lower expected number of units short.

The major difference between foam-based and RFID systems is minimal operating cost. Infrastructure cost and variable cost are the cost factors that differ significantly. Foam-based systems are assumed to operate at lower infrastructure cost than RFID systems and without variable cost. Therefore, foam-based systems show lower overall system costs. In addition, comparing the findings for expected error rates from III.2.3 and III.3.7 suggest that a foam-based system may operate with lower error rates than a RFID system. Lower error rates for foam-based systems would further increase the difference in operating costs.

The computational results for optimal policies of foam-based systems are verified with MATLAB simulations. The approach is equivalent to the one for RFID systems described above. The demand rate \( \lambda \) is 10, and each arriving customer reduces the actual inventory by one. Errors for false negatives and false positives are treated separately. The progression of each error is a function of arriving customers and residual measurement errors for the steps remaining until removal of the last item. Random numbers are used to make a decision on whether errors for false negatives or false positives are reduced. If these numbers are larger than the individual residual measurement errors for false negatives and false positives divided by the number of units left on the shelf, the recorded inventory either remains the same (only a reduction in the number of false negatives), is reduced by one (either no reduction in either error, or a reduction for both errors), or reduced by two (for a sole reduction in the number of false positives). Initial values for inventory levels are given by \( S \) for the actual inventory level and by \( S - (S \cdot d_n) + (S \cdot d_p) \) for the recorded inventory level. The number of units short and the time until replenishment are added up and divided by the number of cycles for ten simulation runs with simulation periods of 10,000 consecutive days. The mean values for number of units short and time until replenishment, along with the corresponding standard deviation are presented in Table 18 together with the computational results. The similarities of the findings indicate that the mathematical
model that describes foam-based inventory management systems generates correct results.

Table 18: Shows the calculated values and simulation results for expected time until replenishment and expected number of units short for different \((r,S)\) policies and detection rates of false negatives and positives. Ten simulations are carried out with simulation periods of 10,000 consecutive days.

<table>
<thead>
<tr>
<th>(d_n)</th>
<th>(d_p)</th>
<th>((r,S))</th>
<th>(E[t_r])</th>
<th>(\mu) of (t_r)</th>
<th>(\sigma) of (t_r)</th>
<th>(E[y^-])</th>
<th>(\mu) of (y^-)</th>
<th>(\sigma) of (y^-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>(1,10)</td>
<td>0.90000</td>
<td>0.89943</td>
<td>0.00291</td>
<td>0.01873</td>
<td>0.01856</td>
<td>0.00108</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>(1,11)</td>
<td>1.00000</td>
<td>0.99962</td>
<td>0.00407</td>
<td>0.01873</td>
<td>0.01863</td>
<td>0.00183</td>
</tr>
<tr>
<td>0.0</td>
<td>0.1</td>
<td>(1,10)</td>
<td>0.91000</td>
<td>0.90829</td>
<td>0.00253</td>
<td>0.03686</td>
<td>0.03693</td>
<td>0.00145</td>
</tr>
<tr>
<td>0.0</td>
<td>0.1</td>
<td>(1,11)</td>
<td>1.00909</td>
<td>1.00802</td>
<td>0.00285</td>
<td>0.03521</td>
<td>0.03505</td>
<td>0.00225</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0</td>
<td>(1,10)</td>
<td>0.88000</td>
<td>0.87935</td>
<td>0.00244</td>
<td>0.01873</td>
<td>0.01863</td>
<td>0.00183</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>(1,10)</td>
<td>0.98182</td>
<td>0.98191</td>
<td>0.00428</td>
<td>0.01873</td>
<td>0.01863</td>
<td>0.00183</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>(1,11)</td>
<td>0.89300</td>
<td>0.89235</td>
<td>0.00357</td>
<td>0.03686</td>
<td>0.03693</td>
<td>0.00145</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>(1,11)</td>
<td>0.99334</td>
<td>0.99195</td>
<td>0.00331</td>
<td>0.03521</td>
<td>0.03505</td>
<td>0.00225</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0</td>
<td>(1,10)</td>
<td>0.85556</td>
<td>0.85643</td>
<td>0.00277</td>
<td>0.03521</td>
<td>0.03505</td>
<td>0.00225</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0</td>
<td>(1,11)</td>
<td>0.96000</td>
<td>0.96023</td>
<td>0.00308</td>
<td>0.03521</td>
<td>0.03505</td>
<td>0.00225</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>(1,10)</td>
<td>0.87267</td>
<td>0.87229</td>
<td>0.00334</td>
<td>0.03521</td>
<td>0.03505</td>
<td>0.00225</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>(1,11)</td>
<td>0.97488</td>
<td>0.97408</td>
<td>0.00386</td>
<td>0.03521</td>
<td>0.03505</td>
<td>0.00225</td>
</tr>
</tbody>
</table>

Note, a \((1,10)\) replenishment strategy shows slightly smaller expected values for units short than a \((1,11)\) replenishment strategy for identical detection rates of false negatives and false positives. This is due to the fact that for the same number of initial errors for systems with different \(S\), the likelihood for an error to have remained in the system when the recorded inventory reaches the threshold is higher for a system with lower \(S\). Consequently, the system with a lower \(S\) is likely to trigger replenishment earlier than the system with a higher \(S\). Early triggering of replenishment reduces the probability for stock-outs.

IV.4.3.2 The effect of higher demand

For the three systems considered in this thesis, higher demand results in higher system cost. (An increase in demand also results in an increase in sales and, consequently, an increase in profit if a positive margin is assumed). The effect of higher demand on operating cost for foam-based systems shows a similar behavior as for RFID system. The levels for threshold and allocated shelf space increase with
increasing demand. Table 19 and Table 20 illustrate how policies are adjusted in order to account for higher demand rates.

**Table 19:** Shows the optimal policy for different detection rates at demand rate $\lambda = 20$ as well as minimal cost, units short per cycle, and corresponding service level.

<table>
<thead>
<tr>
<th>$d_n$</th>
<th>$d_p$</th>
<th>$C_{φ(r,S)}$</th>
<th>Optimal $(r,S)$ policy</th>
<th>$E[y^*]$</th>
<th>$S_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>3444.59</td>
<td>(2,15)</td>
<td>0.00877</td>
<td>0.99971</td>
</tr>
<tr>
<td>0.0</td>
<td>0.1</td>
<td>3472.29</td>
<td>(2,15)</td>
<td>0.01698</td>
<td>0.99943</td>
</tr>
<tr>
<td>0.0</td>
<td>0.2</td>
<td>3486.76</td>
<td>(3,16)</td>
<td>0.00626</td>
<td>0.99978</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0</td>
<td>3461.31</td>
<td>(2,15)</td>
<td>0.00718</td>
<td>0.99976</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>3482.14</td>
<td>(2,15)</td>
<td>0.01461</td>
<td>0.99951</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
<td>3511.72</td>
<td>(3,15)</td>
<td>0.00569</td>
<td>0.99982</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0</td>
<td>3497.28</td>
<td>(2,16)</td>
<td>0.00478</td>
<td>0.99984</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>3511.39</td>
<td>(2,16)</td>
<td>0.01584</td>
<td>0.99944</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>3519.42</td>
<td>(2,16)</td>
<td>0.02101</td>
<td>0.99925</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0</td>
<td>3550.86</td>
<td>(2,17)</td>
<td>0.00314</td>
<td>0.99989</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>3542.64</td>
<td>(2,16)</td>
<td>0.01145</td>
<td>0.99961</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
<td>3538.73</td>
<td>(2,16)</td>
<td>0.01564</td>
<td>0.99946</td>
</tr>
</tbody>
</table>

**Table 20:** Shows the optimal policy for different detection rates at demand rate $\lambda = 30$ as well as minimal cost, units short per cycle, and corresponding service level.

<table>
<thead>
<tr>
<th>$d_n$</th>
<th>$d_p$</th>
<th>$C_{φ(r,S)}$</th>
<th>Optimal $(r,S)$ policy</th>
<th>$E[y^*]$</th>
<th>$S_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>4236.45</td>
<td>(3,19)</td>
<td>0.00877</td>
<td>0.99971</td>
</tr>
<tr>
<td>0.0</td>
<td>0.1</td>
<td>4249.67</td>
<td>(3,19)</td>
<td>0.01698</td>
<td>0.99943</td>
</tr>
<tr>
<td>0.0</td>
<td>0.2</td>
<td>4274.83</td>
<td>(4,19)</td>
<td>0.00626</td>
<td>0.99978</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0</td>
<td>4261.20</td>
<td>(2,19)</td>
<td>0.00718</td>
<td>0.99976</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>4274.63</td>
<td>(3,20)</td>
<td>0.01461</td>
<td>0.99951</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
<td>4284.26</td>
<td>(3,19)</td>
<td>0.00569</td>
<td>0.99982</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0</td>
<td>4273.26</td>
<td>(2,19)</td>
<td>0.00478</td>
<td>0.99984</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>4307.64</td>
<td>(3,20)</td>
<td>0.01584</td>
<td>0.99944</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>4301.40</td>
<td>(3,20)</td>
<td>0.02101</td>
<td>0.99925</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0</td>
<td>4300.68</td>
<td>(2,20)</td>
<td>0.00314</td>
<td>0.99989</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>4348.75</td>
<td>(2,20)</td>
<td>0.01145</td>
<td>0.99961</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
<td>4330.11</td>
<td>(3,20)</td>
<td>0.01564</td>
<td>0.99946</td>
</tr>
</tbody>
</table>
**IV.4.3.3 Sensitivity analysis**

The performance of the foam-based system significantly depends on the input factors. These input factors are equivalent to the ones for RFID systems with the exception of variable costs (e.g. tag cost) that do not apply for foam-based systems. To provide an understanding on the system’s sensitivity to input variation, the cost factors, \( c_s, c_{rep}, \) and \( K_i \) are varied. The effect of such variations on minimal operating cost is examined by plotting these parameters against the output. The relation between input and output variations is found to be approximately linear. The slope coefficients of these plots provide a rough estimate on the impact a change of a parameter has on the system’s performance. For the plot (Figure 65 in B.2):

- Minimal cost versus \( c_{rep} \), a slope coefficient of approximately 340 is derived for variations of \( c_{rep} \) between 0 and 9.
- Minimal cost versus \( \lambda \), a slope coefficient of approximately 60 is derived for variations of \( \lambda \) between 10 and 30.
- Minimal cost versus \( K_i \), a slope coefficient of approximately 10 is derived for variations of \( K_i \) between 5 and 25.
- Minimal cost versus \( c_s \), a slope coefficient of approximately 9 is derived for variations of \( c_s \) between 50 and 350.

The findings are very similar to the findings for RFID systems, and lead to the conclusion that cost factors with higher slope coefficients must be derived more carefully.

Penalty cost shows again a highly non-linear relation to operating cost. Figure 50 illustrates the increase in minimal cost as a response to an increase in penalty cost. Note that the increase in minimal operating cost with increasing penalty cost is smaller than for RFID systems. (For a comparison of Figure 50 and Figure 47, note the difference in the scale for minimal cost). The increase in penalty cost also affects the expected number of units short. The number of expected units short drops with growing penalty cost (Figure 51).
Figure 50: Shows the increase in minimal operating cost as the penalty cost increases (illustrated for different detection rates of false negatives and positives).

Figure 51: Shows the reduction in units short with increasing penalty cost (illustrated for different detection rates of false negatives and positives).
IV.4.3.4 Overall performance analysis

Based on the findings for the operation of periodic review, RFID, and foam-based systems, the performance of each model is compared to the others. The number of variations to inputs is limited in order to provide meaningful results. For this comparison, the cost factors $c_{rep} = $3, $c_s = $110, and $c_p = $10 as well as $t_l = 0.02$ days, $T_r = 1$ day, and $T_h = 365$ days are the same for all three models. In addition to these input factors, there are costs specific to the inventory management policy. The review cost for periodic review is $c_r = $0.5. The tag cost for the RFID system is assumed $0.1 while the infrastructure cost $K_I$ is $12. The infrastructure cost for foam-based systems is $K_I = $7.5.

RFID systems are expected to operate at high read rates. Therefore, read rates of 80%, 90%, and 100% are considered for further analysis. The knowledge on detection rates for false negatives and positives for foam-based systems is limited. In order to avoid neglecting relevant system input information, long ranges for detection rates of 0% to 40% for false negatives and 0% to 20% for false positives are considered. Figure 52 illustrates the minimal operating costs for the periodic review, RFID, and foam-based system for a demand rate of $\lambda = 10$ and the system parameters given above. According to Figure 52, the RFID system with tag cost of $0.1 operates at approximately 12% lower costs than the periodic review system. The foam-based system operates at about 23% lower operating costs than periodic review and at approximately 13% lower costs than RFID with tag cost of $0.1 (Table 21).
Figure 52: Shows a comparison of minimal operating cost for inventory management systems based on periodic review, RFID, and weight-sensitive foam (illustrated for different read rates and detection rates for false negatives and positives).

The differences in minimal operating costs between foam-based systems and other inventory management strategies become more significant for higher demand rates. Arrows in Figure 53 indicate the increasing difference with higher demand rates between minimal operating cost for foam-based systems and RFID systems with tag costs of $0.1. At a demand rate of $\lambda = 20$, the foam-based system shows minimal operating cost of $3482$ ($d_n = 0.1, d_p = 0.1$). For a RFID system, the operating costs are $4242$ for tag cost of $0.1$ and $4972$ for tag cost of $0.2$ (read rate $= 100\%$). The cost for operating a periodic review system is $4690.1$. Consequently, the foam-based system operates at 26% lower costs than a periodic review system and at 18% lower costs than a RFID system (Table 21). For a demand rate of $\lambda = 30$, operating a foam-based system requires $4274.6$, operating a RFID system requires $5417$ and $6512$, respectively, while the operating cost for an inventory management system based on periodic review requires $6066.5$. At this demand rate, replacing the periodic review system with a foam-based system results in a reduction in operating cost of 30%, while a RFID system with tag cost of $0.1$ results in an improvement of 11% (Table 21). A RFID system with tag cost of $0.2$ lead to higher operating cost than periodic review. This is a consequence of the variable cost arising from tag cost that becomes more significant with higher demand. In case the limit for
maximum allocated shelf space is 20 items, a foam-based inventory system shows a reduction in operating costs of 67% for $\lambda = 20$, and of 90% for $\lambda = 30$. A RFID system with tag cost of $0.1$ leads to a reduction of 60% for $\lambda = 20$ and 87% for $\lambda = 30$ for the case of limited shelf space.

Figure 53: Shows a comparison of minimal operating cost for inventory management systems based on periodic review, RFID, and weight-sensitive foam for $\lambda = 10$, 20, and 30 (illustrated for different read rates and detection rates for false negatives and positives).

Table 21: Shows expected improvements in operating costs for the replacement of one system by another one (estimations for $\phi=1.0$, $d_n = 0.1$, and $d_p = 0.1$).

<table>
<thead>
<tr>
<th>Expected improvements in operating cost of</th>
<th>Foam over periodic review</th>
<th>RFID over periodic review</th>
<th>Foam over RFID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c_{tag} = $0.1 $c_{tag} = $0.2 $c_{tag} = $0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 10$</td>
<td>23%</td>
<td>12%</td>
<td>1%</td>
</tr>
<tr>
<td>$\lambda = 20$</td>
<td>26%</td>
<td>10%</td>
<td>-6%</td>
</tr>
<tr>
<td>$\lambda = 30$</td>
<td>30%</td>
<td>11%</td>
<td>-7%</td>
</tr>
</tbody>
</table>

The performance analysis provided above assumes that a foam-based system is operated with an infrastructure cost of $7.5$ per item per year. However, there is limited data available to support this assumption. Figure 54, Figure 55, and Figure 56 show that even with an increase in infrastructure cost, a foam-based system still
operates at lower costs than a RFID system. The maximum for the infrastructure cost at which both systems show equal operating costs are given by the cross-over points in Figure 54, Figure 55, and Figure 56. The intersection of minimal cost for a foam-based system with $d_n = 0.2$ and $d_p = 0.1$ and RFID with tag cost of $0.1$ form the lower limit at $40$ for a demand rate of $\lambda = 10$ (Table 22). The upper limit is given with $90$ by the intersection of the minimal operating cost for a foam-based system with $d_n = 0$ and $d_p = 0$ and RFID with tag cost of $0.2$. For a demand rate of $\lambda = 20$, the lower limit moves to $50$ while the upper limit increases to $120$, and for a demand rate of $\lambda = 30$, the lower limit is $70$ and the upper limit becomes $150$. Consequently, weight-sensitive foam may show a significantly higher infrastructure cost than RFID while operating at lower inventory management cost. This is due to the fact that weight-sensitive foam systems do not introduce variable costs such as tag costs.

Table 22: Gives the infrastructure costs for foam-based systems at which these systems operate at lower cost than a comparable RFID system.

<table>
<thead>
<tr>
<th></th>
<th>Foam</th>
<th>RFID</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = 10$</td>
<td>$d_n = 0$, $d_p = 0$</td>
<td>$K_{I_{Foam}} &lt; 40$</td>
</tr>
<tr>
<td></td>
<td>$d_n = 0.2$, $d_p = 0.1$</td>
<td>$K_{I_{Foam}} &lt; 40$</td>
</tr>
<tr>
<td>$\lambda = 20$</td>
<td>$d_n = 0$, $d_p = 0$</td>
<td>$K_{I_{Foam}} &lt; 60$</td>
</tr>
<tr>
<td></td>
<td>$d_n = 0.2$, $d_p = 0.1$</td>
<td>$K_{I_{Foam}} &lt; 50$</td>
</tr>
<tr>
<td>$\lambda = 30$</td>
<td>$d_n = 0$, $d_p = 0$</td>
<td>$K_{I_{Foam}} &lt; 70$</td>
</tr>
<tr>
<td></td>
<td>$d_n = 0.2$, $d_h = 0.1$</td>
<td>$K_{I_{Foam}} &lt; 70$</td>
</tr>
</tbody>
</table>
Figure 54: Shows the minimal cost for foam-based systems for different infrastructure cost in reference to RFID systems with tag costs of $0.1 and $0.2, respectively (demand rate $\lambda = 20$).

Figure 55: Shows the minimal cost for foam-based systems for different infrastructure cost in reference to RFID systems with tag costs of $0.1 and $0.2, respectively (demand rate $\lambda = 20$).
Figure 56: Shows the minimal cost for foam-based systems for different infrastructure cost in reference to RFID systems with tag costs of $0.1 and $0.2, respectively (demand rate $\lambda = 30$).
This chapter summarizes the key findings of the thesis and discusses critical aspects thereof. In addition, it outlines the implications of the findings and distinguishes between theoretical and practical implications. The chapter ends with a discussion of future prospects.
V.1 Key Findings

The core business of the retail industry is to breakdown bulk products into smaller lot-sizes or individual items and sell the items to the general public at a convenient location to consumers. The highly competitive retail industry operates with high sales volumes and low margins. In order to establish a profitable and sustainable position against its competitors, a retailer is required to operate at low cost for the customer’s benefit while providing high product availability.

This thesis is concerned with the optimization of retail store shelf inventory management with respect to minimal process costs. Currently, response-based in-store logistics are widely established in the industry. Store clerks visually inspect inventory levels on retail shelves by regularly walking the aisles. They add products to picking lists for replenishments from the backroom if inventory levels are low. The economical effectiveness of such a periodic review policy is influenced by various factors. Among the most important factors are labor cost that arises through the manual inspection of inventory levels and accuracy of data collection. A consequence of high labor cost is often a low observation frequency that may cause delays in the detection of low inventory levels.

This thesis outlined that operational logistics cost and on-hand inventory are two main resources to reach a certain service level. These resources form the major success factors of the retail business together with on-shelf product availability. A major logistics operation within the retail store is the replenishment process, which is triggered by the detection of low shelf stocks. Retailers significantly rely on inventory management systems to track historical demand patterns, monitor inventory levels, and to determine optimal safety stock levels that balance product availability and inventory holding cost. However, despite the necessity of accurate information for efficient inventory management, retailers fail to accurately keep track of sold merchandise and on-shelf inventory levels. This failure derives from the fact that data generated by the current electronic systems are highly inaccurate. Inaccurate information about on-shelf inventory may result in untimely and non-optimal decisions for replenishment. Poor shelf inventory management may cause many stock-out situations. Studies indicate that stock-out rates at retail stores are around 8.2%. Stock-out situations result in a diminution of several percentages of sales which lead to a reduction in profit of up to 10%. Additionally, retailers lose about 13% of their customers due to stock-out situations. Seventy-two percent of the root causes for stock-out situations are traced to the retail store. Store shelving and store ordering constitute the most important store related causes of stock-outs. It was
found that inventory inaccuracy significantly contributes to stock-outs that are caused by improper store shelving and store ordering. Reasons for inventory inaccuracy are found in scanning process errors, misplacements, improper handling of returns, theft, damages, fraud, and other process errors. Some retailers try to improve their stores’ data accuracy through periodic inventory audits. These audits are significantly labor intensive and current audit periods of six months to a year are still too long to effectively eliminate errors from the inventory system.

**What technologies show potential to support shelf stock monitoring?**

Controllability and observability are of major concern for each system and may be realized in many different ways. However, improvements in controllability and observability are characterized by a positive effect on major retail business success factors, including logistics operation cost or product availability. Consequently, suggestions for improvements are critically compared to the performance of processes and policies currently established. This thesis investigated whether the improvement of on-shelf inventory accuracy through high fidelity stock monitoring on retail shelves lowers overall inventory management costs. Specifically, this thesis examined the potentials of RFID at item level that induces fix infrastructure cost as well as variable tag cost accruing for each item. The potential benefits arising from RFID are compared to those accredited to the weight-sensitive foam specifically designed for the shelf stock monitoring application. In contrast to RFID, weight-sensitive foam does not introduce variable sensor cost for object detection. Therefore, weight-sensitive foam suggests lower overall operations cost.

**What are the properties of these technologies?**

Applying RFID to shelf stock monitoring assumes that each item on display is equipped with a RFID tag. Stock quantities are detected through RFID antennae that are incorporated into the shelves. Shelf inventory levels are updated continuously and used for replenishment decision making. This approach seeks to overcome inaccurate estimations of shelf stocks caused by erroneous point-of-sale data and infrequent visual observation by store clerks. Retrieving on-shelf inventory data directly by measuring the number of displayed items significantly improves inventory data accuracy, which is essential for informed management decisions. However, studies indicate that read rates for RFID tags may often not reach 100%. Consequently, the technology used to improve the quality of inventory data may introduce measurement errors itself. In order to provide a fair estimate on the potential of RFID to improve data accuracy and lower operating costs, the limitations associated with this technology have to be taken into account. This thesis
specifically addressed measurement errors that occur from shielding. Shielding refers to the event where an item or its RFID tag blocks the RFID reader’s signal and hence prevents the reader from detecting another RFID tag and its associated item. Consequently, the recorded inventory system may conclude stock quantities that are too low. However, the measurement error is dynamic as items become detectable when those items are removed that caused the shielding. This thesis developed mathematical models that account for the imperfect state information resulting from measurement errors. These models provide an assessment of the economical effect of on-shelf inventory monitoring technology on the inventory management cost.

RFID allows for remote detection and identification of objects. It shows potential for various applications in the retail sector and other industries. Although the capability of remote detection significantly improves logistics processes along the supply chain, this capability is not a necessity for on-shelf inventory management. This thesis developed weight-sensitive foam designed specifically for product monitoring on retail shelves. In an attempt to meet the retail industry’s requirements for very low cost, the foam-based inventory management system uses packaging foam that usually finds applications in the automotive and construction industries due to its lightweight structure, excellent thermal characteristics, and low manufacturing costs. Weight-sensitive foam directly detects objects through their physical characteristics instead of through an intermediate device such as a RFID tag. On a retail shelf, products are displayed on top of such foam where the foam is used as a weight-receptive spacer material for capacitive sensor elements. The exposure to weight leads to local deflections of the foam. These deflections reduce the gaps between electrodes of capacitive sensor elements and result in changes in capacitance. These changes in capacitance are directly related to the weights of the products. By separating individual items, the inventory management system concludes the total stock quantities.

This thesis examined the properties of polymer foam to serve as weight-receptive material. The stress-strain relationship of such polymer foam was found to be nearly linear for moderate stress levels of up to 0.1MPa, but nonlinearities and hysteresis are distinctive for higher stress levels. For pressure of up to 0.04MPa, an elastic modulus of approximately 4.4MPa was derived that puts the weight of a product and the corresponding measurement of foam deflection in an explicit relation. In addition, it was found that for the application of weight detection, foam may only be used at low pressure levels where it shows the characteristics of linear elastic solid.
For higher stress levels, the elastic modulus was found highly nonlinear indicating a softening and densification region. In the linear region, the elastic modulus of the foam was found more than two orders of magnitude smaller than the bulk modulus of polymers. This result was unexpected because a large number of closed-cell foams show linear scaling of the relative elastic modulus versus relative density. A cross-section image of an electron microscope scan revealed that the foam consists of only few large anisotropically-shaped voids in the foam’s cross-section. These voids were identified to account for the high elasticity in the foam’s cross-section because the elastic modulus of the foam significantly depends on the bending characteristics of the two-dimensional walls of the cells. The scaling laws derived from small and isotropically-distributed spherical void were found not applicable.

The foam’s weight sensitivity was determined to be less than 10g/cm². This sensitivity is sufficient for the detection of a large variety of retail products. A prototype was developed to test the design and to retrieve information about measurement accuracies. These tests revealed that detection errors occur when completely- or partly-covered sensors do not report a load (false negatives) or when sensors show the detection of a product when they are not completely loaded (false positives). False positives only occur due to poor alignment of objects. However, changes in capacitance due to cross-talk (from a loaded sensor to an unloaded one) are below the noise level of 5fF. Preliminary measurements found error rates for false negatives ranging from 1% to 9.7%, and error rates for false positives ranging from 0% to 4%. It was also observed that both errors become zero with the removal of all items.

In order to examine the potential of weight-sensitive foam to serve as shelf stock monitoring technology, errors in object detection were taken into account. In contrast to RFID’s susceptibility to false negatives, weight-sensitive foam may also record false positives. The errors may not be easily discernible because the effects of false positives and false negatives on the recorded inventory level may partly compensate one another. The mathematical model presented in this thesis accounts for these errors in order to provide an accurate evaluation of the economic potential arising from introducing weight-sensitive foam into the on-shelf inventory management concept. Furthermore, a detailed cost analysis demonstrates that the proposed sensor network may be manufactured at $0.9 per running meter, and that the total system cost including a data processing unit, communication module, and batteries may be manufactured at about $10 per running meter. The infrastructure cost arising from the deployment of weight-sensitive foam is also considered as one of the input parameters to the system.
**How can a mathematical model of an on-shelf inventory system account for imperfect state information?**

This thesis evaluated new possibilities for more sophisticated and efficient control of shelf inventory by using on-shelf inventory monitoring technology to reduce operations cost. In order to evaluate these effects, this thesis developed mathematical models that describe the cost factors of the management process and account for the imperfect operation of monitoring technology. The management systems presented in this thesis are concerned with balancing shelf allocation cost, inventory review cost, replenishment cost, costs for the operation of information processing systems, and stock-out cost. The on-shelf inventory decreases with arriving customers. The customer inter-arrival times were modeled according to a Poisson process. The inventory level is reviewed either periodically in the case of manual inspection, or continuously if the management system makes use of shelf stock monitoring technology. If the recorded inventory level is equal to or lower than the threshold at any review point, the actual inventory level is raised up to the number of items for which shelf space is allocated. This procedure is in contrast to the general understanding of inventory systems with positive lead times where the actual inventory level increases after a lead time by the amount that was ordered at the review point. Usually, the ordered amount is the difference between the base-stock and the recorded inventory level at the review point. With the occurrence of demand during lead time, the inventory level is lower than the base-stock level after replenishment. Therefore, these inventory management problems have to be treated as multi-period problems. However, it was argued that retail shelves are commonly replenished from boxes and containers up to the base-stock level as opposed to moving only a numbered amount of products from the backroom to the shelves. Super-numerous items are returned to the backroom for storage. With this understanding of the shelf replenishment process, the on-shelf inventory management problem is modeled as a single period problem where each period begins with identical inventory levels. The actual inventory level is equivalent to the base-stock level while the recorded inventory differs from the actual inventory by the measurement errors between the two. With each removal of an item from the shelf the difference between actual and recorded inventory may change depending on whether or not that item caused a measurement error recently eliminated from the system. Because measurement errors are caused by items on display, RFID measurement errors will have disappeared from the system with the removal of the second to last item on display. In contrast, weight-sensitive foam errors will have disappeared from the system with the removal of the last item on display. Therefore, the measurement error changes over a period of occurrences of demand that is equal
to the base-stock level, according to an arbitrary sequence. With each occurrence of demand, the measurement errors are either reduced or remain the same. For each number of demand occurrences, the probability was derived that the measurement errors have been reduced by a certain amount. Based on these probabilities, the expected time until replenishment, the expected cycle time, the expected number of units short, and the expected cost per planning horizon were calculated. The optimal control parameters were determined by the set of $r$ and $S$ that minimizes the cost function.

**How do optimized $(r,S)$ replenishment policies for different technologies compare?**

Single-period models allow for a simple analysis of expected inventory management costs. The control parameters that minimize the expected inventory management costs constitute the optimal replenishment policy. The minimal operating costs for periodic review, RFID, and foam-based inventory management systems were identified for different demand rates, varying input cost factors, different read rates and tag costs for RFID systems as well as different error detection rates of false negatives and false positives for foam-based systems. It was shown that at low demand rates, RFID systems operate at lower cost than periodic review only for high read rates and low tag cost of $0.1. The reduction in operating cost is approximately 10%-12% in comparison to an ordinary periodic review strategy and independent of the demand rate. In comparison to periodic review, a RFID system with tag cost of $0.2 performs better only for low demand rates, but shows significantly higher operating costs for higher demand rates. An on-shelf inventory management system based on weight-sensitive foam may lower operating cost by 23% for low demand rates and by 26% to 30% for higher demand rates in comparison to the operating costs associated with a periodic review system. In comparison to RFID systems with tag costs of $0.1, a foam-based system may lower operating cost by 13% to 22%. These findings are based on infrastructure cost of $7.5 per allocated shelf space per item per year for a foam-based system. This estimate of infrastructure cost is subject to uncertainty. However, this seems to be a conservative estimate which makes the findings on operations cost meaningful. It was observed that foam-based systems show lower operating costs than RFID systems. This is true as long as the infrastructure cost of the foam-based system remains below $40 per allocated shelf space per item per year in the case of low demand and low RFID tag cost. For high demand, infrastructure cost may be increased up to $150 per allocated shelf space per item per year and still yield to lower operations cost. A major difference in operating cost between RFID and
foam-based systems is constituted by the variable tag costs. Therefore, a foam-based system shows lower on-shelf inventory management cost than a RFID system even if the infrastructure cost of the foam-based system is higher.

The customer service level was found to remain at a very high level in the case of unlimited allocated shelf space. The control parameters “allocated shelf space” and “threshold level” are adjusted to compensate for high demand. The analysis of optimal control parameters for different inventory management systems revealed that the control parameters for continuous monitoring are significantly lower than those for periodic review. This finding has significant practical implications for situations where shelf space is limited (or where opportunity costs increase excessively). As the values of the control parameters increase with increasing demand, the base-stock level for periodic review systems eventually reaches the limit for allocated shelf space. If the control parameters are restricted in adjusting for high demand due to shelf space limitation, the customer service level drops significantly with increasing demand and stock-outs occur more frequently. The base-stock level of a periodic review system reaches the limit for allocated shelf space at a lower demand rate than the base-stock level of a continuous review system. Therefore, the impact on operating costs due to penalty costs arising from stock-outs is more significant for periodic review systems than for continuous review systems.

The individual answers to the research sub-questions provide a thorough understanding of how automatic stock monitoring on retail shelves can improve shelf inventory management with respect to minimal total process costs.

V.2 Discussion of Research Findings

V.2.1 Operations costs

The findings in this thesis indicate that the costs for on-shelf inventory management and shelf replenishment can be significantly reduced by replacing the currently established periodic review inventory management policy with a management system based on weight-sensitive foam. Foam-based systems operate at significantly lower cost than comparable RFID-based systems due to the fact that foam-based systems introduce lower infrastructure costs and no variable cost. However, the exact infrastructure cost associated with a foam-based system is subject to uncertainty because current estimates are solely based on findings derived
from a prototype. The actual implementation and operation of such a system may reveal that current assumptions for infrastructure cost are inaccurate, which may require a reevaluation of the system’s operations cost. In addition, infrastructure costs for RFID systems are also subject to uncertainty and may be lower than $12 per allocated shelf per item per year as assumed for the model presented in this thesis. Nevertheless, the results in the previous subsections indicate that even with a significant increase in infrastructure cost of foam-based systems, these systems still outperform periodic review and RFID systems.

In order to reduce RFID infrastructure cost and to increase the accuracy of location information, this thesis introduced a distance-sensitive high frequency RFID system. Current RFID antennae are non-directional in order to cover a wide angle. To provide detailed information about the location of products measured with respect to their distance to the antenna, the RF signal strength of such an antenna may be reduced to limit the antenna field. However, a limited read range requires the installation of more antennae. This thesis demonstrated the feasibility of distance-sensitive high frequency RFID systems where RFID antennae operate at high signal strengths to cover the widest area possible. These antennae are still capable of providing location information of products through the detection of the products’ distances to the antenna. However, the analysis of RFID on-shelf inventory did not examine the economical effects that may result from a reduction in infrastructure cost due to fewer antennae in combination with an increase in tag cost due to more sophisticated tags.

The scope of this thesis is limited to on-shelf inventory management. Therefore, the models account for costs arising from stock monitoring and replenishment. Profits are assumed to occur through sales that over-compensate the costs related to managing the on-shelf inventory. Limiting the scope to on-shelf inventory management allows for a thorough investigation of all cost factors relevant to shelf stock management. However, this limitation neglects further implications of the technology for other areas of store and retail supply chain management. For example, if RFID tagging at item level had penetrated the entire supply chain and RFID infrastructure was already in place at the retail store, the costs for RFID infrastructure and tags used for the cost analysis would have to be adjusted. Basically, such a scenario would significantly reduce the operating costs for RFID systems. Indeed, the costs incurred by RFID could be shared with other applications such as supply chain management, faster check-outs, etc., where benefits arise through the use of RFID.
An on-shelf inventory system based on weight-sensitive foam shows an advantage over a RFID system constituted by the fact that a foam-based system may be easily installed at only specific locations at the retail store, e.g. promotional shelves. The products on sale do not have to be modified. This advantage represents a major difference when compared to RFID systems (unless RFID has already been deployed area-wide). In order to provide on-shelf inventory monitoring with RFID, either the manufacturer or the distributor has to equip the products that are on sale with RFID tags for the limited time period of the promotion. This requirement includes the need for sophisticated logistics and significant investments in technology to specifically equip a very limited number of items with RFID tags, rendering the use of RFID economically inefficient.

The comparison of periodic review models with continuous on-shelf inventory monitoring already revealed that the requirements for maximum allocated shelf space are significantly lower for continuous review models. However, the analysis of operating costs has not yet included benefits, other than reduced costs for allocated shelf space, that may arise from shelf space that frees up due to lower safety stocks. A retailer may achieve additional sales due to a higher product variety or may cut down sales floor infrastructure cost because the same amount of products is offered on a smaller shop floor.

V.2.2 Implications of a higher inventory management abstraction level

This thesis assumes that a retailer is specifically interested in precise information about on-shelf inventory in order to make informed decisions regarding an inventory management strategy. However, some retailers may prefer to manage their inventories at a higher abstraction level such as sales floor inventory or retail store inventory. These retailers may still benefit from RFID at case level where a gate-reader between the backroom and the sales floor keeps track of replenishments. It would not be necessary to record specific quantities that are moved from the backroom onto the sales floor, because the detection of a replenishment event would indicate that the shelf has been completely restocked. With each event of replenishment the actual on-shelf inventory level would be raised to the allocated shelf space and the recorded inventory level would be adjusted accordingly. Therefore, the detection of replenishment would be considered equivalent to the detection of a replenished shelf.

However, these retailers renounce detailed information about on-shelf inventory at specific locations on the sales floor. It may be difficult and cumbersome to avoid shelf stock-out situations at large retail stores where identical products are displayed
at different shelf locations throughout the store. Indeed, the on-shelf inventory management for promotional items may also be affected significantly. During a sales promotion, products are often displayed at additional, prominent locations in the store such as the entrance or the check-out counter. The information provided by a sales floor inventory management system may be misleading because promotional shelves may deplete while the inventory level still shows remaining stocks. However, these products may be on display at other locations in the store. Consequently, the store manager would not be informed about a depleted promotional shelf and the promotion would not take effect.

V.2.3 Imperfect state information due to shrinkage and misplacements

The models developed in this thesis aimed to provide a better understanding of the dynamics of an on-shelf inventory system susceptible to measurement errors of monitoring technology, but do not address other sources of errors that may distort the recorded inventory level. These sources are primarily shrinkage and misplacements. (Another primary source for errors is transaction errors. However, these errors are not considered here because they have no immediate effect on shelf inventory). Shrinkage is defined as theft (theft by employees and shop lifting) and unsaleables (damaged, perished or expired goods) that often occurs unnoticed and results in discrepancies between actual and recorded inventory. While shrinkage may occur at each stage of the retail supply chain, its occurrence at the retail store may significantly affect on-shelf product availability. Shrinkage does not only distort the inventory records but it also affects current replenishment policies. Review periods that are assumed to meet certain demand forecasts may show poor performance when the inventory levels decrease more rapidly due to shrinkage.

Misplacements refer to events where items are removed from their designated location on a shelf and arbitrarily put back on a shelf at another location in the store. Although these misplaced products are still on the sales floor and the product count appears correctly in the inventory records, misplaced items will not be found by customers or store clerks. Consequently, these items will not be sold until they have been located and returned to the shelf location assigned to that product category. Due to misplacements, inventory records may not correctly reflect the stock quantities available at a product’s designated shelf location. This may lead to a reduction in product availability.

An on-shelf inventory management system that uses RFID at item level in order to keep a record of stock quantities on display would treat shrinkage as additional demand and update the inventory records immediately after shrinkage occurs. The occurrence of shrinkage would become visible if the inventory records are matched
with point-of-sales data. A RFID-based system would also be capable of detecting misplacements (if they are not shielded by other products) through the unique ID that would allow identifying the corresponding product category. Store clerks could be navigated to the specific location of the misplaced product in order to return it to its designated shelf location.

A foam-based on-shelf inventory system also shows capabilities to adjust its recorded inventory levels to the occurrences of shrinkage. However, misplacements are more difficult to detect. Weight-sensitive foam distinguishes individual products through their pressure patterns in order to conclude the number of items on display. While this approach allows the count of items, it often fails to correctly identify an item’s product category. The reason is that the packaging of many products shows similar or identical characteristics (e.g. cans and bottles for different products often show identical shapes and weight patterns). However, for products with distinctive differences in pressure patterns, inventory systems based on weight-receptive foam may detect misplacements.

Operating a RFID gate-reader between the backroom and the sales floor in order to keep a record of sales floor inventory operates at a level of abstraction that does not allow for the detection of shrinkage or misplacements. However, if shelves are restocked up to the allocated shelf space during replenishment, discrepancies in recorded inventory due to shrinkage or misplacements are compensated with each replenishment cycle. This is because the recorded inventory levels are reset to the base-stock level after replenishment. Therefore, shrinkage or misplacements do not significantly affect product availability on retail shelves as long as they occur at a low rate.

**V.2.4 Concurrent operation of RFID and weight-sensitive foam**

This thesis argued that leading retailers introduce RFID into their supply chains at case and pallet levels for automatic product identification and tracking. Currently, retailers are hesitant to promote wide-area applications at item level because the potential economic benefits are unclear. It was also argued that shelf stock monitoring based on RFID is likely to require a large scale implementation in order to become economically feasible. In addition, isolated applications such as equipping promotional goods with RFID tags appear impractical. Consequently, RFID and weight-sensitive foam may be applied concurrently. The retail supply chain could benefit from RFID at case level to track and trace cases and pallets along the supply chain, and in-store operations could benefit from detailed shelf stock information due to weight-sensitive foam. In this case, the high variable tag
costs for item-level tagging would not apply and consumer privacy concerns due to RFID would not arise.

V.3 Theoretical Implications

V.3.1 Physics

In the area of soft matter physics, this thesis contributed to the theory with a detailed analysis of elastic properties of polyolefin foam. This thesis identified the distinctive regions of polyolefin foam that characterize its elastic properties and identified the elastic modulus of foam. The elastic modulus was found more than two orders of magnitude smaller than the bulk modulus of polymers. This result was unexpected because a large number of closed-cell foams show linear scaling of the relative elastic modulus versus relative density. This thesis identified that the few anisotropically-shaped voids in the foam’s cross-section and the bending characteristics of their two-dimensional cell walls constitute the reason for a low elastic modulus.

V.3.2 Operations research

This thesis contributes to the existing theory in operations research with a theoretical framework that conceptualizes the effect of shelf stock monitoring technology on shelf inventory management. Previous research had developed frameworks to examine the effects of identification technology on inventory data accuracy but neglected to account for measurement errors introduced by the monitoring technology itself. This thesis investigated the errors that may be introduced to the system through measurement errors and analyzed the system’s behavior with respect to the stochastic nature of such errors. The result is mathematical models that are not limited to the assumption of perfect operation of monitoring technology. Instead, the mathematical models developed in this thesis allow optimizing inventory management systems under imperfect state information. Incorporating imperfect state information into an assessment of the economic impact on shelf inventory management costs provides a more sophisticated understanding of actual benefits.
V.4 Practical Implications

The highly competitive nature of the retail business requires retailers to continuously increase the efficiency of their logistics processes. At the retail store level, retailers may gain an advantage through careful control of on-shelf inventory. This thesis finds that inventory management costs are lowered significantly if manual, periodic review is replaced with continuous review of inventory levels at low review cost. The cost improvements that may arise from such technology considerably depend on the costs associated with current in-store logistics processes and expenses for deployment, operation, and maintenance of shelf stock monitoring technology. This thesis illustrated the dimension of potential benefits that may be achieved with a continuous review strategy. However, the input parameters to the control system vary from retailer to retailer (e.g. labor cost, shelf space allocation cost, etc.) and so vary the potential benefits. Retailers will need to conduct their own analysis with respect to input parameters specific to their business. The framework developed in this thesis along with the mathematical models will support retailers with evaluating the cost-efficiency of their current replenishment policies and with deriving explicit results on potential benefits for their retail stores. The sensitivity analyses for the different models identify the input factors that should receive most attention during the data gathering process that precedes the system evaluation.

This thesis demonstrates that technologies with potential for improvements of on-shelf inventory management are not limited to RFID. In fact, the findings reveal that a replacement of periodic review for RFID systems with tag cost of $0.2 does not seem feasible unless RFID generates economic value in areas other than on-shelf inventory management. Under the assumption of effective operation, the introduction of RFID should be considered if:

- it becomes a necessity to comply with mandates or regulations.
- tag costs are low.
- allocated shelf space is strongly limited and periodic review policies result in significant stock-out rates.
- substantial benefits can be created in other areas such as supply chain logistics, faster check-out, etc.

However, other technologies may allow reducing in-store operating costs more significantly than RFID. The development of weight-sensitive foam and its economic effects on shelf inventory management are presented as an alternative to RFID. Retailers who do not intend to introduce RFID on a large scale may still consider deploying weight-sensitive foam on their shelves that experience high
demand (e.g. promotional shelves.)

The mathematical models that were used to assess the potential economic benefits arising from the introduction of shelf stock monitoring allow retailers to adjust the input parameters and verify the findings specific to the measurement errors that they experience. Measurement errors and input parameters may vary significantly from retailer to retailer as infrastructure setup, hardware, and products change. By using the mathematical models developed in this thesis, retailers may realize economic benefits through automatic shelf stock monitoring despite the presence of measurement errors.

In addition, the mathematical models permit for fast calculations of optimal replenishment strategies as opposed to simulations. Therefore, an inventory management system may recalculate and readjust an optimal replenishment strategy to a change observed in input parameters, e.g. a change in demand rate.

Beyond lowering operation costs and providing high product availability, high fidelity shelf stock monitoring may lower overall shelf stocks as safety stocks become obsolete with accurate detection of shelf depletion and timely replenishment. The monetary benefits gained from reducing safety stocks and freeing up capital may lead to investments in other areas to generate higher profits.

V.5 Future Perspectives

The scope of this thesis is limited to on-shelf inventory management and the effects of high fidelity shelf stock monitoring on operations costs. The cause for out-of-stock situations on retail shelves is assumed to be an unawareness of shelf stock depletion in combination with untimely replenishment from the backroom. The backroom inventory level is assumed to always be sufficiently high to allow for complete shelf replenishments. This thesis focused on shelf inventory management and considered backroom inventory as part of an overall effort to manage retail store inventory levels. The limitation to on-shelf inventory management allows for a detailed analysis of the effects of automatic stock monitoring on inventory management costs. However, it is desirable to take backroom inventory levels into consideration because they may experience stock-out situations. Stock-outs in the backroom directly affect on-shelf inventory management.

In addition, this thesis assumed that the review time for periodic review inventory management strategies and the replenishment lead time remained constant over time. This assumption seems appropriate for the analysis of the effects of shelf stock
monitoring technology on shelf inventory management. However, review and replenishment lead time may show a certain degree of variation. These times may be affected by the work load of store clerks and may vary over the day or the week. Accounting for such variation increases the complexity of the mathematical models, but may provide more accurate results on actual operations costs.

Furthermore, this thesis addressed the notion that shrinkage and misplacements are other primary sources of error that distort the recorded inventory level. These sources of error were not accounted for in the mathematical models in order to obtain an explicit relation between variations in measurement accuracy and changes to the system outputs. However, the current analysis should be extended to allow for inventory distortions due to shrinkage and misplacements. For example, the effects of shrinkage on actual inventory may be modeled similar to demand according to a Poisson process with a shrinkage rate $\mu$ that is significantly smaller than demand rate $\lambda$. The downside of such an extension to the current models is that it makes them more complex, and the analysis of best replenishment policies more cumbersome. Simulations may provide a simple means to obtain a fundamental understanding of how shrinkage and misplacements affect optimal system control parameters and operations costs without having to deal with complex mathematical relations.

This thesis stated that the introduction of distance-sensitive high frequency RFID systems may lower RFID infrastructure cost while increasing tag costs. Additional benefits may arise through more accurate location information of products and the ability of location clustering. The current mathematical models do not account for such benefits. Whether such modifications to the existing technology increase economic benefits in inventory management remains to be evaluated.

This thesis aimed to create awareness on how high fidelity stock monitoring for retail shelves may improve on-shelf inventory management. This thesis evaluated and tested different technologies to provide information on detection accuracies for shelf inventory levels. The mathematical models developed to assess the economic impact of high fidelity stock monitoring on the operation of on-shelf inventory management indicated that retailers may significantly benefit from the introduction of such technology. The development and installation of prototypes will serve to verify the accuracy of the research findings. In turn, these prototypes should reveal further potential challenges that did not receive enough consideration. Based on the gravimetric approach for product detection on retail shelves presented in this thesis, two European companies are currently working on a prototype for field testing.
Field tests will not only increase the level of confidence on the potential economic benefits assessed in this thesis, but it will also help in finding optimal grid sizes for the electrode leads on foam. The design presented in this thesis proved adequate for the limited number of products evaluated. However, it is not clear whether the same grid size is feasible for all product categories or whether foam with different grid sizes will be necessary.
Appendices
Appendix A

A.1 Weight-sensitivity measurements for numerous foam

In order to choose polyolefin foam to serve as mount for the electrodes, the responsibilities to applied forces were tested for numerous foams. While high receptiveness is desirable to provide sufficient sensitivity to detect even light products, the foam must still feature high resilience and provide a stable stance to the products on display. The foams tested were selected to represent a broad range of polyolefin foam available. These foams are distinguishable by slight modifications of the physical properties. The tested foams are TEE0300.46, TEE0400.20, TEE0400.20LT, TEE0400.48W, and S604.25a.

The applied force was linearly increased and decreased from 0N to 1N, which corresponds to a variation from 0g/cm² to 100g/cm² in 3 consecutive cycles. The force was applied by a Zwick Z0.5 (basic line) with a maximum applied force of 500N. The changes in capacitances was measured with an AD7745 Sigma-Delta converter from Analog Devices and recorded on a PC. The results are shown in Figure 57, Figure 58, Figure 59, Figure 60, and Figure 61.

The TEE0300.46 shows high resilience, but when compared to the TEE0300.25 in Figure 34 the TEE0300.46 shows only a low change in capacitance of 15fF. The TEE0400.20 shows a nonlinear and irreproducible change in capacitance, which makes this foam unfeasible for the dedicated application. The TEE0400.20LT shows excellent resilience, but the changes in capacitances of maximum 35fF are smaller than the 50fF derived from the TEE0300.25. The TEE0400.48W shows fluctuations in the change of capacitance that result in moderate resilience. Additionally, the maximum change in capacitance of 30fF is only moderate. Alveo’s S604.25a shows absolutely irreproducible results. Hence, the Alveo S604.25a is not feasible for the retail application. The poor results may be affected less by the elastic properties of the foam than by the poor leads. Due to the porous surface of this foam, the leads forming the electrodes were poorly painted.
Figure 57: Alveo’s Tee 0300.46.
Figure 58: Alveo’s Tee 0400.20.
Figure 59: Alveo’s TEE0400.20LT.
Figure 60: Alveo’s TEE0400.48W.
Figure 61: Alveo’s S604.25a.
Appendix B

B.1 Equations

Proof of Eq. (37):

\[ P(\xi_k = n) = \binom{k}{n} \cdot \frac{(\Delta_0)_n \cdot (S-1-\Delta_0)}{(S-1)_k} \]

\[ = \frac{k!}{n!(k-n)!} \cdot \frac{(\Delta_0)_n \cdot (S-1-\Delta_0)}{(S-1)_k} \]

\[ = \frac{(\Delta_0)_n \cdot (S-1-\Delta_0)}{k!} \cdot \frac{(k-n)!}{(S-1)_k} \]

\[ = \frac{(\Delta_0)_n}{n} \cdot \binom{S-1-\Delta_0}{k-n} \]

\[ = \frac{\Delta_0}{n} \cdot \binom{S-1}{k} \]

\[ \square \]
Complete form of Eq. (68)

\[
\begin{align*}
E[f_r] &= \sum_{k=1}^{S} \lambda(\text{prob}(S_{-}(-\Delta_{p0} + \Delta_{m0}) - k + (\xi_{p_{k-1}} + \xi_{m_{k-1}}) = r + 1)) \\
&+ \text{prob}(S_{-}(-\Delta_{p0} + \Delta_{m0}) - (k-1) + (\xi_{p_{k-1}} + \xi_{m_{k-1}}) = r + 2) \\
&+ \text{prob}(S_{-}(-\Delta_{p0} + \Delta_{m0}) - (k-1) + (\xi_{p_{k-1}} + \xi_{m_{k-1}}) = r + 1) \\
&+ \text{prob}(S_{-}(-\Delta_{p0} + \Delta_{m0}) - (k-1) + (\xi_{p_{k-1}} + \xi_{m_{k-1}}) = r + 2)
\end{align*}
\]
Complete form of Eq. (68) (continued)

\[
X = \left( \begin{array}{c}
(S - \Delta_{n_0}) (k - 1) \\
S - (k - 1)
\end{array} \right) + \left( \begin{array}{c}
\Delta_{n_0}^{-i} \\
S - (k - 1)
\end{array} \right) \cdot \left( \begin{array}{c}
\Delta_{n_0}^{-i} \\
S - (k - 1)
\end{array} \right)
\]

\[
y = \frac{S - \Delta_{n_0}^{-i}}{S - (k - 1)} \\
E[v] = \sum_{k=1}^{K} \frac{S}{k} \sum_{i} x_i
\]
B.2 Simulation results

Table 23: Shows the simulation results for mean time until replenishment and units short for two sets of control parameters. Four simulations are carried out with a simulation period of 25,000 consecutive days each.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Run 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,11)</td>
<td>0.98951</td>
<td>0.98937</td>
<td>0.98956</td>
<td>0.99237</td>
<td>0.01617</td>
<td>0.01627</td>
<td>0.01769</td>
<td>0.01823</td>
</tr>
<tr>
<td>(2,11)</td>
<td>0.87965</td>
<td>0.87761</td>
<td>0.88048</td>
<td>0.87823</td>
<td>0.00076</td>
<td>0.00128</td>
<td>0.00068</td>
<td>0.00079</td>
</tr>
</tbody>
</table>

Figure 62: Shows the minimal costs for RFID inventory systems with tag cost of $0.2 and different demand rates ($\lambda = 10, 20, 30$).
Figure 63: Shows the plots for $c_s$ and $p_c$ versus minimal cost and service level versus $p_c$.

Figure 64: Shows the different plots for $K_I$, $c_r$, $c_s$, and $\lambda$ versus minimal cost. For reasons of readability, only read rates of 100% and 90% are shown. The plots for lower read rates, however, show similar shapes.
Figure 65: Shows the different plots for KI, c_r, c_s, and λ versus minimal cost. For reasons of readability, only d_n = 0.1 and d_p = 0 are shown. The plots for lower read rates, however, show similar shapes.
References


156. EPCglobal, *EPCglobal Tag Data Standards Version 1.3*, 2006.


Curriculum Vitae

Personal Data

Date of Birth:  October, 28, 1978  
Birthplace:  Zurich, Switzerland  
Citizenship:  Swiss

Education

ETH (Swiss Federal Institute of Technology), Zurich, Switzerland  2005-2008
PhD candidate in Elgar Fleisch’s Information Management Group, Department of Management, Technology, and Economics.
Research focus: Operations Management

MIT (Massachusetts Institute of Technology), Cambridge, MA, USA  2007
Visiting scientist in Stanley Gershwin’s Group, Department of Mechanical Engineering.
Research focus: Operations research, logistics process optimization

Georgia Institute of Technology, Atlanta, GA, USA  2004
Visiting scientist and student supervisor in Thad Starner’s Group, College of Computing.
Research focus: Contextual computing, wearable computing, HCI

ETH (Swiss Federal Institute of Technology), Zurich, Switzerland  1998-2004
MSc in Electrical Engineering and Information Technology