A Review of Holonic Manufacturing Systems Literature

Authored by Amro M. Farid

Supervisor: Dr. Duncan McFarlane

Center for Distributed Automation & Control
Institute for Manufacturing
Engineering Department
University of Cambridge

June 9, 2004
Abstract

This reports gives an extensive review of the holonic manufacturing systems literature in 2004. After a brief introduction of holonic principles and concepts, a full rationale of holonic systems in manufacturing is given. In particular, an argument is developed that holonic manufacturing systems provide greater agility than their current industrially adopted counterparts. The body of the review highlights the most significant architectures, methodologies, protocols, algorithms and interactions. Finally, open research issues and barriers to industrial adoption are extracted from the review of the current state of the literature.
## Contents

1 Introduction ................................................................. 7
   1.1 Background to Holonic Systems ................................. 8
   1.2 Holonic Systems in Manufacturing Context .................. 9

2 HMS Rationale & Its Requirements .................................. 12
   2.1 Business Drivers Rationale ................................. 12
   2.2 Manufacturing System Rationale .............................. 13
      2.2.1 CIM Systems .................................. 14
      2.2.2 Heterarchical Systems ............................. 18
      2.2.3 Holonic Systems ................................. 19
   2.3 Control System ................................................. 21
      2.3.1 Requirements and Recommendations .................... 21
      2.3.2 Available Technologies .............................. 22

3 Developments in Holonic Manufacturing Systems .............. 23
   3.1 Architectures ................................................... 24
      3.1.1 Conceptual Description .............................. 24
         3.1.1.1 Initial Conceptions .................. 24
         3.1.1.2 Cooperation Domains .................. 25
         3.1.1.3 PROSA – Product, Resource, Order, Staff Architecture .......... 26
         3.1.1.4 HCBA – Holonic Component Based Architecture ........... 27
         3.1.1.5 HoMuCS – Holonic Multi-Cell Control System .......... 29
         3.1.1.6 Agent Architecture Taxonomy .................. 29
            3.1.1.6.1 Reactive Agents .................. 30
            3.1.1.6.2 Deliberative Agents .......... 31
            3.1.1.6.3 Hybrid Agents ............... 31
         3.1.1.7 Multi-Agent System Taxonomy .................. 32
            3.1.1.7.1 Federation Architectures .... 33
            3.1.1.7.2 Autonomous Agent Architectures .... 34
         3.1.1.8 Miscellaneous Conceptions .................. 34
      3.1.2 Detailed Design and Simulated Implementations .......... 34
         3.1.2.1 Rockwell/BHP Steel Rod Mill Water Cooling System .... 35
         3.1.2.2 VTT Automation Robot Cell ......... 35
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.2.3</td>
<td>Holden’s Engine Operations Engine Assembly Line</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>3.1.2.4</td>
<td>Miscellaneous Designs</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>3.1.3</td>
<td>Implementations</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>3.1.3.1</td>
<td>Daimler Chrysler Holomobiles</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>3.1.3.2</td>
<td>University of Cambridge Robot Assembly Cell</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>3.1.3.3</td>
<td>KU Leuven</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>3.1.3.4</td>
<td>Miscellaneous Implementations</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>Methodologies &amp; Strategies</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>3.2.1</td>
<td>Design &amp; Modelling Tools</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>3.2.1.1</td>
<td>IDEF0</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>3.2.1.2</td>
<td>Petri-Nets</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>3.2.1.3</td>
<td>The Unified Modelling Language – UML</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>3.2.1.4</td>
<td>Agent Building Tools</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>3.2.2</td>
<td>Design Methodologies &amp; Strategies</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>3.2.2.1</td>
<td>PROSA</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>3.2.2.2</td>
<td>HoMuCS Methodology</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>3.2.2.3</td>
<td>DACS Multi-Agent System Design Methodology</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>3.2.3</td>
<td>Migration Methodologies &amp; Strategies</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>3.2.3.1</td>
<td>HCBA Migration Strategy</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>3.2.3.2</td>
<td>Daimler-Chrysler Holomobiles Migration Strategy</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>3.2.3.3</td>
<td>Holden’s Engine Operations Migration Strategy</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>Protocols</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>3.3.1</td>
<td>Contract-Net Protocol</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>3.3.1.1</td>
<td>The Tri-Base Acquaintance Model</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>3.4</td>
<td>Algorithms</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>3.4.1</td>
<td>Scheduling Algorithms</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>3.4.1.1</td>
<td>Lagrangian Relaxation Scheduling</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>3.4.1.2</td>
<td>Temperature Equilibrium Scheduling</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>3.4.1.3</td>
<td>Market-Driven Scheduling</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>3.4.2</td>
<td>Integrated Planning Scheduling Algorithms</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>3.4.2.1</td>
<td>Lagrangian Relaxation with Planning</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>3.4.2.2</td>
<td>Market Based Approaches</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>Interactions</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>3.5.1</td>
<td>IEC 61499 Function Blocks</td>
<td>59</td>
<td></td>
</tr>
</tbody>
</table>
3.5.2 Function Block Operating System ................................. 61

4 Open Issues & Industrial Adoption ................................. 62

4.1 HMS System Design .................................................. 62

4.1.1 Availability of Proven Design Methodologies .................. 62

4.1.2 Analysis of Performance of Holonic Manufacturing Systems 63

4.2 HMS System Implementation ......................................... 63

4.2.1 Appropriate environments for implementing holonic control . 63

4.2.2 Establishing suitable standards for holonic control systems . 63
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A Generic Architecture of a Holon [42]</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>A Waterfall Chart Linking Business Drivers to Control System Needs [175][44]</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>The Basic Functions of Manufacturing Facility [199]</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>A Typical Manufacturing Control Hierarchy [172]</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>An Abstracted Hierarchical Structure [77], [29]</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>An Abstracted Heterarchical Structure [28], [29]</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>An Abstracted Holarchy Structure [28][29]</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>A General Activity Model of a Holon [61]</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>The Critical Holon Interfaces in Holonic Manufacturing Systems [61]</td>
<td>25</td>
</tr>
<tr>
<td>10</td>
<td>Basic Buildings Blocks of PROSA Architecture [246]</td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>Aggregation and Specialization in PROSA Workshop [245]</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>HCBA Holon Architecture [54]</td>
<td>28</td>
</tr>
<tr>
<td>13</td>
<td>HCBA System Architecture [58]</td>
<td>29</td>
</tr>
<tr>
<td>14</td>
<td>Outline of HoMuCS Architecture [138]</td>
<td>30</td>
</tr>
<tr>
<td>15</td>
<td>A Subsumption Agent: A Reactive Architecture [43][40]</td>
<td>31</td>
</tr>
<tr>
<td>16</td>
<td>A BDI Agent: A Deliberative Architecture [43][197]</td>
<td>32</td>
</tr>
<tr>
<td>17</td>
<td>An InteRRaP Agent: A Hybrid Architecture [43][177]</td>
<td>32</td>
</tr>
<tr>
<td>18</td>
<td>A Desktop Utility Agent System: A Facilitator Architecture [65]</td>
<td>33</td>
</tr>
<tr>
<td>19</td>
<td>An Example Broker Architecture [211]</td>
<td>33</td>
</tr>
<tr>
<td>20</td>
<td>An Example Mediator Architecture [211]</td>
<td>34</td>
</tr>
<tr>
<td>21</td>
<td>A Conceptualized Autonomous Agent Architecture [211]</td>
<td>35</td>
</tr>
<tr>
<td>23</td>
<td>Holon Architecture [4] [170]</td>
<td>36</td>
</tr>
<tr>
<td>24</td>
<td>The VTT Automation Holon Structure [114]</td>
<td>38</td>
</tr>
<tr>
<td>25</td>
<td>The VTT Simulated Manufacturing Robot Cell [114]</td>
<td>39</td>
</tr>
<tr>
<td>26</td>
<td>The Flat VTT Robot Cell Holarchy [114]</td>
<td>39</td>
</tr>
<tr>
<td>27</td>
<td>A Conventionally Controlled Mercedes Benz V6/V8 Engine Assembly Line [47]</td>
<td>41</td>
</tr>
<tr>
<td>28</td>
<td>A Conventionally Controlled Line with Implemented Holons [47]</td>
<td>41</td>
</tr>
<tr>
<td>29</td>
<td>A Holonically Assembled Meter Box [55][259]</td>
<td>42</td>
</tr>
<tr>
<td>30</td>
<td>Robotic Assembly Cell Meter Box Diagram [55][259]</td>
<td>43</td>
</tr>
<tr>
<td>31</td>
<td>University of Cambridge Holonic Robot Assembly Cell [55][259]</td>
<td>43</td>
</tr>
<tr>
<td>32</td>
<td>Static and Dynamic HCBA Integration [55]</td>
<td>44</td>
</tr>
<tr>
<td>33</td>
<td>University of Cambridge Robot Assembly Cell HCBA Architecture [55]</td>
<td>44</td>
</tr>
<tr>
<td>34</td>
<td>Holonic Flexible Assembly System at KU Leuven [241]</td>
<td>45</td>
</tr>
<tr>
<td>35</td>
<td>An Illustrative IDEF0 Example</td>
<td>47</td>
</tr>
<tr>
<td>36</td>
<td>An Illustrative Petri-Nets Example [199]</td>
<td>48</td>
</tr>
<tr>
<td>37</td>
<td>An Illustrative PROSA based UML example [246]</td>
<td>48</td>
</tr>
<tr>
<td>38</td>
<td>An Overview of the HoMuCS Methodology [138]</td>
<td>50</td>
</tr>
<tr>
<td>39</td>
<td>A Migration Framework for a Holonic Factory [57]</td>
<td>52</td>
</tr>
<tr>
<td>40</td>
<td>The Contract-Net Protocol [43][224]</td>
<td>54</td>
</tr>
<tr>
<td>41</td>
<td>The IEC 61499 Function Block Model [39]</td>
<td>60</td>
</tr>
<tr>
<td>42</td>
<td>IEC 61499 Function Block Data/Event Synchronization [39]</td>
<td>61</td>
</tr>
<tr>
<td>43</td>
<td>A FBOS based Holonic Manufacturing System</td>
<td>62</td>
</tr>
</tbody>
</table>
1 Introduction

The Holonic Manufacturing Systems (HMS) field was initiated in Japan by Suda [230], [231] as a response to the growing perception that Japanese manufacturing firms lacked competitiveness in a global manufacturing environment. He hypothesized that the cause of this inability to compete was rigid manufacturing practices that did not have the necessary agility \(^1\) and responsiveness \(^2\) in increasingly volatile markets. Suda noted that the robustness \(^3\), flexibility \(^4\), and adaptability \(^5\) of holonic systems would be a highly desirable characteristic in the effort to regain international competitiveness \(^6\).

Since that point, numerous and relevant technology contributions have been made in the field of holonic manufacturing systems. This review paper summarizes the most relevant and

\(^1\) Definition 1.0.1. Agility – The quickness with which a system adapts to modifications of the production and/or product envelop in the sense of operation flexibility \([100]\). It is both reactive (responsive) and proactive \([208]\).

\(^2\) Definition 1.0.2. Responsiveness – The ability of a production system to respond to dynamic conditions (originating inside or outside the manufacturing organization) which impact upon production goals \([160]\). It forms the reactive part of agility \([208]\).

\(^3\) Definition 1.0.3. Robustness – The ability of a system to not have to respond or change its behaviour because it is designed well enough so that these disturbances do not affect the output \([100]\).

\(^4\) Definition 1.0.4. Flexibility – The facility with which a system design may be modified to meet similar product requirements \([100]\), \([67]\) and extended with new elements to augment the existing level of functionality \([223]\), \([67]\), \([209]\).

\(^5\) Definition 1.0.5. Adaptability – The ability of a system in operation to change behaviour to maintain its desired output in the presence of external and/or internal disturbances \([100]\).

\(^6\) Robustness, flexibility and adaptability are necessary parts of responsive and agile systems. The first being a subset of the second \([208]\).
effective of these contributions with the intention of identifying the features, attributes and modules that would lead most effectively to an industrial realization of a holonic manufacturing system. In doing so, it relies upon previously published reviews [44], [172], [175] and builds upon them wherever possible.

This report illustrates these points in the following sequential method of approach. This introduction continues in Section 1.1 with a description of holonic systems’ origin, and later in Section 1.2 gives a very topical description of how holonic principles may be applied in manufacturing. Section 2 presents a three-part holonic manufacturing system rationale that builds from the business drivers to the manufacturing system implications and their requirements on the chosen control system. Section 3 forms the body of this review and consists of the technological developments most relevant to holonic manufacturing systems. These developments can be divided into: architectures which vary in detail from conceptual descriptions to partial implementations, design methodologies which attempt to specify systematic ways to develop detailed architectures, and finally the protocols, algorithms, and interactions which together give holonic manufacturing systems their required agility. Finally, Section 4 closes the review by identifying the open issues, specifically in HMS design and implementation that prevent their straightforward industrial adoption.

1.1 Background to Holonic Systems

Holonic systems originate from the works of the philosopher A. Koestler [134], who in 1967 proposed the term holon to describe his observations of the behaviour of biological and social systems. The word holon itself originates from the Greek word "holos" meaning "whole" and the suffix "on" meaning "part of" i.e. a neutron or proton. He found that all biological and social systems evolve, grow, and adapt to complex and changing environments by forming stable intermediate holons. More specifically, holons exhibit a dual behaviour which he called the Janus Effect [38]. On the one hand, each holon has an autonomous quality. Its development and functionality is sufficient to exist alone. On the other hand, each holon also has a cooperative quality that allows it to depend upon a social framework of holons. In such a way, they interact together to meet overall goals of the collective [134].

It was these ideas which Suda felt would be particularly beneficial in a manufacturing system [230], [231]. Soon afterwards, the Holonic Manufacturing Systems Project with its associated research consortium was formed as one of the six Intelligent Manufacturing Systems (IMS) feasibility studies [112], [207], [100], [64]. Since then, numerous conceptions (of
varying degrees of similarity) were proposed to bring holonic principles to a manufacturing context. In order to instantiate more concretely such a holon, one generally accepted conceptual architecture is introduced in Figure 1. Its composition is explained in further detail in Section 1.2.

![Figure 1: A Generic Architecture of a Holon [42]](image)

1.2 Holonic Systems in Manufacturing Context

The holon concept above must adhere rigorously to holonic behaviours, properties and attributes which have been carefully defined by the HMS consortium. It is defined below.

**Definition 1.2.1.** Holon – An autonomous and cooperative building block of a manufacturing system for transforming, transporting, storing and/or validating information and physical objects. A holon consists of an information processing part and often a physical part. A holon can be part of another holon [100].

In addition to the properties of autonomy and cooperation, some authors [158], [175], [44] add recursivity, self-organization, and reconfigurability to the list of manufacturing holon properties. These properties, upon which the holon definition rests, are also defined below.

**Definition 1.2.2.** Autonomy – The capability of an entity to create and control the execution of its own plans and/or strategies [100].
**Definition 1.2.3.** Cooperation – A process whereby a set of entities develops mutually acceptable plans and executes these plans [100].

**Definition 1.2.4.** Recursivity – A similarity in the informational architecture and communications model between holons [158].

**Definition 1.2.5.** Self-Organization – The ability of manufacturing units to collect and arrange themselves in order to achieve a production goal [100].

**Definition 1.2.6.** Reconfigurability – The ability of a function of a manufacturing unit to be simply altered in a timely and cost effective manner [100].

One should note that the last three properties directly contribute to the definitions of robustness, flexibility and adaptability found on page 7. Additionally, the holonic literature often refers to holonic attributes, and for clarity, its definition is given below.

**Definition 1.2.7.** Holonic Attributes: Attributes of an entity that make it a holon. The minimum set is autonomy and cooperation [100].

Finally, to clarify holonic systems in a manufacturing context, the definition of a holonic manufacturing system is given with that of a holarchy of which HMS are composed.

**Definition 1.2.8.** Holonic Manufacturing System: A holarchy that integrates the entire range of manufacturing activities from order booking through design, production and marketing to realize the agile manufacturing enterprise [100].

**Definition 1.2.9.** Holarchy: A system of holons that can cooperate to achieve a goal or objective. Tho holarchy defines the basic rules for cooperation for the holons and thereby limits their autonomy [100].

In the previous section, an initial conception of a holon was given in Figure 1. Having defined the necessary set of behavioral properties, attention can now turn to the elements of its composition which would allow it to achieve these behaviours. A holon is necessarily composed of a physical, hardware part and a decision-making software part which are connected by an intra-holon interface [64]. Additionally, each holon has both a holon human interface and an inter-holon interface for communication directed to achieving global objectives [42].

The physical hardware part of a holon is traditionally thought of as the manufacturing resources which one may find in a plant such as a milling or stamping machines [100].

---

7Throughout this review, the author will refer to holonic attributes and properties interchangeably.
Their control might be a combination of programmable logic controllers (PLC), local area networks (LAN), and PC’s [100]. However, this conception is not sufficient. Firstly, not any machine can be readily included into a holon architecture. Specifically, it needs to be sufficiently flexible [8] such that 1.) it may interface with the software component of the holon and 2.) meet the previously specified behavioural requirements [100]. Good examples of flexible manufacturing resources are CNC milling machines and robots [100]. The above conception is also insufficient because it does not include the raw material, works in progress (WIP), and final products to which a software interface maybe added to achieve holonic behaviour [270], [174], [41]. Finally, the above conception is insufficient because it does not account for the ability of a holon to exist within another holon like cells, shop floors, or factory holons [245].

The HMS consortium has almost universally accepted that the software part of a holon and its holarchy is enabled by agents and multi-agent systems [24]. Brennan [39] explains this relationship between holons and agents in great detail and concludes that multi-agent systems are a necessary component of any HMS implementation. However, there exist many (not necessarily congruent) ideas of the definition of an agent [9]. The author resolves the apparent disagreement with a definition formed from Huhns [122] [10] and Durfee [83] [11] which is found below:

**Definition 1.2.10.** Multi-Agent Systems – Loosely-coupled network(s) of active and persistent software components that perceive, reason and communicate together to solve problems that are beyond their individual capabilities.

The reader accustomed to object-oriented programming may note that agents are similar to objects in their focus on data abstraction, encapsulation, modularity and inheritance [33]. However, they do differ in that they possess a greater degree of autonomy to control their state and behavior, can demonstrate reactive, proactive, and cooperative behavior and have their own thread of control [260].

---

[8] How flexible a manufacturing resource in the holonic context is still an open research question, but the implemented architectures in Section 3.1.3 should shed light onto the degree of required flexibility.

[9] For example, Rzevski defines agents as "a software object that mimics the role of a competent personal assistant to perform a specific task on behalf of a user intelligently or not, independently or with little guidance [200]."

[10] Huhns states that multi-agents systems are "a group of active, persistent software components that perceive, reason, and communicate."

[11] Durfee states that multi-agents systems are "loosely-coupled network(s) of problem solvers that work together to solve problems that are beyond their individual capabilities."
This section has briefly described the necessary behaviour and composition of a generic holon. From this platform, Section 2 can give a rationale for the behaviour of holonic manufacturing systems and Section 3 can describe the technological developments of which it is composed. This section, however, has not given sufficient attention to the specific functionality of an individual holon or holonic manufacturing system. In later sections, it will be important to identify clearly the functionality of the proposed systems. This given, this review restricts its discussion strictly to discrete part manufacturing despite the clear potential use of holonic and multi-agent manufacturing systems to Human-Machine Information Management [1], [2], Robotics [8], Cognitive Psychology [268], Automated Storage and Retrieval Systems (warehousing)[143], Product Design [66], [15] [69], [187], Enterprise Integration and Supply Chain Management [97], [185], [192], and Process Control Systems [60].

2 HMS Rationale & Its Requirements

The previous section introduced and defined holonic concepts within the domain of discrete part manufacturing. It also superficially stated the rationale for its emergence. This section describes that rationale in much greater detail. It uses a three part method to highlight how the holonic vision directly links the salient business drivers to the necessary manufacturing system attributes which in turn are used to develop the implied functionality requirements of its control system. The basic outline of this rationale is summarized graphically in the waterfall chart in Figure 2 below.

2.1 Business Drivers Rationale

Manufacturing firms find themselves in a continually evolving and increasingly competitive marketplace. In addition to consistently providing high quality products at low cost, they must realize that their excess capacity causes their marketing departments to gradually shift market power to the consumer [44]. As a result, to maintain market share, firms must address consumer demands for continual innovation, low-cost customization, and improved customer service [207]. These demands are satisfied by shortened product-life cycles, reduced time-to-market, and increased product variety at little or no extra cost or quality deterioration.  

12 The author believes that many sources in the literature do not explicitly contrast the observed functionality of a proposed system to the functional requirements derived from the previously stated holonic behaviours.
Figure 2: A Waterfall Chart Linking Business Drivers to Control System Needs[175][44]

Their fulfillment manifests themselves in the firm’s manufacturing operations as 1.) more complex products, 2.) rapidly evolving product lines 3.) faster product introductions 4.) volatile output (& demand) and 5.) reduced investment per unit [175]. In short, the manufacturing firms must find ways to mitigate the increased complexity and continual change in its operations [44].

### 2.2 Manufacturing System Rationale

The difficulty in addressing the previously mentioned operations requirements can not be well understood without a clear understanding of the nature of modern-day industrially implemented manufacturing systems. These systems may be broadly categorized under the title of computer integrated manufacturing (CIM). Their limitations and drawbacks provide the motivation from which HMS research has emerged. The defining characteristics of these CIM systems are now presented so that later in Sections 2.2.2 and 2.2.3 we may understand the motivation behind the innovations and research that have followed.
2.2.1 CIM Systems

Much like the 1990’s business environment outlined in Section 2.1, the 1980’s experienced its own competitive drivers. Firms were under particular pressure to reduce their time-to-market and operating costs [108]. More specifically, firms found themselves paying excessively for large quantities of inventory and many layers of management personnel [108]. Manufacturing firms sought to achieve these objectives from the state of the art management techniques of the time like Just-In-Time (JIT), Total-Quality-Management (TQM) and World-Class-Management (WCM) [108]. All of these techniques were system-wide improvements and required a great deal of coordinated and organized data collection. With the emergence of cost-effective local area networks and personal computers, CIM systems were able to gain prevalence as the need to implement JIP, TQM, and WCM with integrated IT systems grew [108].

Harrington, the first to introduce CIM, captured its function in its original definition:

**Definition 2.2.1.** Computer Integrated Manufacturing (CIM) – The integration of business engineering, manufacturing and management information that spans company functions from marketing to product distribution [109].

This rather broad definition of CIM led to more reflective acronyms like the Computer Integrated Enterprise (CIE) and the Computer Integrated Manufacturing Enterprise (CIME) [13], but they were not generally accepted [108]. This was partially because the integration of manufacturing operations remained at the core of CIM functionality. Explicitly stated, CIM systems integrated shop floor processes, their manufacturing engineering planning, and the production planning and control of the shop floor and its associated materials [109].

CIM systems achieved their previously mentioned objectives in a number of ways – all tied to the organization and coordination of data through the through networked PC [108].

For the first time, extensive use of the databases managed all of the functional [14], product [15],

---

13 The European Commission sponsored programme ESPRIT endorsed this acronym.

14 Definition 2.2.2. Functional Data – Information categorized as part of a company’s knowledge base [108].

15 Definition 2.2.3. Product Data – Data on products and their products [108].
operational \(^{16}\) and performance \(^{17}\) data of an enterprise \([108]\). This data was published with word processors, manipulated with spreadsheets, and transferred instantaneously through the network \([108]\). CIM also automated communication and data collection within the factory; thereby improving its speed and accuracy \([199]\). CIM, (in many instances), even eliminated the use of paper and its associated cost \([108]\).

This degree of coordination affected the operations and interfaces of every department of the manufacturing enterprise. Marketing, engineering, production planning, plant operations, physical distribution and business & financial management were all improved \([108]\). While not all of these interfaces can be easily identified, Figure 3 systematically captures the dominant interactions between the firm’s constituent departments \([199]\). Manufacturing firms achieved further productivity gains through the use of numerous inter-departmental PC-based software applications and algorithms \([108]\). Most notably, Computer Aided Process Planning (CAPP) sped up the preparation of routing and operation plans and Master Production Scheduling (MPS) planned the mixture of products for manufacture through the consolidation of data on existing and forecasted order quantities \([108]\). Materials Requirement Planning (MRP) managed the purchasing of materials through the extensive use of accurate inventory level information and products’ bill of materials \([199]\). Additionally, Manufacturing Resource Planning (MRPII) created production schedules based upon the available manufacturing capacity \([108]\). Finally, firms were able to promote simultaneous or concurrent engineering through the use of Computer Aided Design and Computer Aided Manufacturing (CAD/CAM) packages \([199]\). The interested reader will find a number of thorough texts on CIM systems \([199]\), \([198]\), \([133]\), \([205]\), \([108]\). The results of this integration and automation led to reductions of design cost of 15-30%, reductions of in-shop part time of 30-60%, quality improvement and scrap reduction of 20-50% and improved product design through an increased number of design iterations \([199]\).

CIM systems provided many benefits with their automated integrative functions but

---

\(^{16}\) Definition 2.2.4. Operational Data – The plans and instructions required to control the operations of a company \([108]\).

\(^{17}\) Definition 2.2.5. Performance Data – Data that measures the degree to which operational or product data has been achieved \([108]\).
proved to have poor agility [240] due to their hierarchical implementations [127]. The most well known CIM architectures proposed by IBM, the National Institutes of Sciences & Technology (NIST), Siemens, Digital Equipment, and the ESPRIT projects were all hierarchical [199], [78], [32]. A generic diagram of CIM hierarchy, seen in Figure 4, shows the generalized manufacturing tasks of planning, scheduling, execution, and control of machines & devices [172]. Each level has its own purpose and function [77]. Higher levels have increasingly complex computing structures and create the global goals and long-range strategies for all the levels (and computers) below them [77] 18. Additionally, sensed information is abstract,

18Interestingly, the design community’s experience and training in the use of hierarchical techniques probably aided the development of CIM as a hierarchical structure [77]. It was probably reinforced by the hierarchical structure of companies as responsibility and accountability are maximized when CIM layer corresponds directly to a personnel department [108]. Nevertheless, recent CIM developments have introduced modified
databases, time period.

Figure 5 show this hierarchical structure as an abstracted tree of parent-child nodes composed of master-slave communication relationships [28]. Commands flow purely top-down and responses return in a static and deterministic fashion after task completion [29].

Beginning with the late 1980’s, some began to think that hierarchical structures were not sufficiently flexible, and adaptable, and fault tolerant [196] [19]. The lack of flexibility results hierarchical structures with increased distribution and coordination [29]. Newer texts on the subject also emphasize integration over hierarchy [108].
from the rigid communication relationships fixed in the early design stages of the CIM system [80]. Flexibility is further degraded when a great deal of complexity is introduced as each controller requires knowledge of its neighbours below and above to maintain these fixed relationships [80]. This complexity is amplified when explicit interrelationship exception handling is written to improve fault-tolerance. Fairley notes that a high degree of modularity and low degree of coupling facilitate implementation, debugging, testing, maintenance and complexity reduction [84]. Complexity also weakens fault-tolerance especially in the presence of disturbances and system modifications [49]. Additionally, CIM systems lack fault-tolerance because they are particularly sensitive to two failure modes. The failure of 1.) the controller or 2.) the communication link below it results in a paralysis of all the controllers to which they are connected [49], [116]. Finally, CIM systems exhibit poor adaptability because neither higher level controllers have access to detailed sensory information nor lower level controllers have the computational resources to make effective use of the data [77]. As a result, data is aggregated and thereby delayed in database formation [77]. In the presence of disturbances, the optimized global decisions of higher level controllers use obsolete or estimated data [77], [196].

This system wide lack of agility of CIM gained American political attention in the early 1990’s [185] much like it had in Japan [230], [231]. As a result, numerous agile manufacturing initiatives outside of holonic manufacturing systems were begun [185]. McFarlane notes some of these initiatives to develop distributed and modular equipment: Reconfigurable Manufacturing (USA), Intelligent Assembly, Fixturing & Handling (Japan,EU) and Self-Diagnosis and Self Repair capabilities (EU/Aus)[171]. Unfortunately, it was soon realized that the implementation of flexible equipment and tooling was not sufficient and a comprehensive system wide view of agility was required [142], [216], [202].

2.2.2 Heterarchical Systems

A system wide approach to agility was proposed with the conception of heterarchical systems [28], [29]. Several researchers, [196], [80], [111], [209], [81], [79], sought to replace the rigidity of hierarchical systems with a completely flat structure where each component exhibits full local autonomy [77]. Decision making is made entirely locally at the point of information

**Definition 2.2.6.** Fault-Tolerance – The ability of a system to continue to function, perhaps in a degraded state, despite the occurrence of system failures [34]. It is a necessary requirement for robustness and adaptability.
gathering and measurement [81]. Furthermore, each component cooperates via a negotiation procedure in the form of temporary and flexible relationships [29]. Figure 6 shows an abstracted view of a heterarchical structure.

Heterarchy had numerous but ultimately insufficient advantages for industrial adoption. The full autonomy of each component in combination with the negotiation algorithm meant that failures did not propagate and the system exhibited fault tolerance [80], [224]. For similar reasons, the system could adapt readily to local disturbances [77]. These results were supported both theoretically [68], as well as experimentally [82]. Additionally, developer found that complexity, measured in lines of code, was reduced [80]. Finally, the system flexibility was excellent as the negotiation algorithm and system components could be modified, removed, and added easily [77]. Heterarchical systems, however, in their full autonomy lost the ability to create and achieve global objectives [80], [24]. Furthermore, Valckenaers showed that the system can reach unstable states where small disturbances induce large disturbances elsewhere in the system [240]. Additionally, hierarchical systems are limited by the network bandwidth as network traffic can grow quickly with the number of components [132]. Finally, heterarchical systems where never industrially adopted due to their inability to achieve a predictable result [28], [29].

2.2.3 Holonic Systems

The failure of heterarchical systems, and the need to resolve CIM systems’ deficiencies opened the way for numerous new manufacturing concepts – loosely categorized as Next Generation Manufacturing Systems. In addition to holonic manufacturing systems, other concepts like bionic [183], [184], genetic [237], [238], fractal [250], random [124], and virtual manufacturing systems [129] were proposed [32], [245]. In addition to the these references, the interested reader can find review and comparisons of these concepts in [244], [235], [236]. Interestingly, a number of authors believe holonic manufacturing to be among the better concepts [101],
Bongaerts and Van Brussel state that holonic systems are preferable because they emphasize autonomous holons that maintain robustness and adaptability in the presence of disturbances, loose hierarchies that permit global optimization, flexible hierarchies that are reconfigurable and adaptable and migration strategies that facilitate industrial adoption [32], [245].

The qualities of autonomous holons and loose-flexible holarctic is a fusion of the advantages of hierarchical and hierarchical systems [32]. Hierarchy is introduced to not just permit global optimization but also to create stability in the system [245] as in the watchmakers parable told in [134], [222]. At the same time, these structural relationship are neither fixed nor permanent. Holons can: belong to multiple hierarchies, enter and leave them and do not rely on the proper function of their neighbours [32]. Figure 7 gives an abstracted view of a holarchy. Additionally, holonic systems have autonomous and cooperative qualities much like heterarchical systems do[44]. They negotiate amongst each other and make local decisions, albeit with less autonomy than their heterarchical system counterparts [245]. More central holons, also have the ability to give advice or apply flexible rules to more peripheral holons [32]. In these ways, holonic manufacturing systems can mitigate most unwanted circumstances and become robust[135], flexible and adaptable [175].

Figure 7: An Abstracted Holarchy Structure [28][29]
2.3 Control System

Having studied the strengths and weakness of hierarchical and heterarchical systems, one can begin to specify more detailed requirements for a holarchy’s behavior – its constituent control system. In this section, the author extracts a number of clearly stated requirements from Sections 2.2.1 and 2.2.2, compares them to recommendations elsewhere in the literature, and finally proposed the available technologies to achieve them.

2.3.1 Requirements and Recommendations

Section 2.2.1 showed that CIM systems were rigid due to fixed master-slave relationships and the complexity between them. From these observations, we can infer two requirements to improve flexibility and robustness:

Requirement 2.3.1. Inter-holon communication relationships must be flexible and temporary.

Requirement 2.3.2. Complexity must be minimized by low-coupling, high cohesion holons.

These requirements are expanded to a more generalized recommendation:

Recommendation 2.3.1. Control interactions should be abstract, generalized and flexible [175].

In other words, the control system must be distributed to achieve a holon’s cooperative, recursive, and reconfigurable properties.

The above requirements also directly improve adaptability. Brennan explicitly states that a distributed systems can more effectively reject disturbances [39]. Adaptability, is furthered by two more requirements:

Requirement 2.3.3. Time sensitive or time varying data must be used locally at the point of gathering or measurement. Time invariant data may be centralized.

Requirement 2.3.4. Time sensitive or time varying data must be available to the relevant decision maker be they local or central.

which may be generalized to the following recommendation.

Recommendation 2.3.2. The architecture of the control should be decentralized and physically-based. [175].
Together, these requirements and recommendations may be used to support Parunak’s assertion that decentralized control must be "emergent rather than planned, concurrent rather than sequential" [186]. Finally, Brennan encourages physically based holons and contrasts a physical decomposition to a functional one [39]. Weiss cites evidence of the former in natural systems where critical information is often organized by physical entities and managed by other physical entities that represent them [258]. Booch finds the latter to cause inconsistency and unintended relationships [33]. One sees that these arguments solidify holons’ autonomous properties within a flexible hierarchy [44].

Finally, Section 2.2.1 showed that CIM systems did not provide enough computation resources to lower level controllers. Hence,

**Requirement 2.3.5.** Lower level controllers must have sufficient computational ability to make their designated decisions.

This requirement can be expanded to two more recommendations:

**Recommendation 2.3.3.** The control should be both reactive and proactive [175].

**Recommendation 2.3.4.** The control should be self-organizing [175].

One may note that this comprises a fundamental upgrade of previously held paradigms of adaptability. Not only must holons respond to disturbances, but they must also self-organize and be proactive towards achieve their global and local objectives. In this way, one may conceive the harmonization of all of the previously mentioned holon properties.

### 2.3.2 Available Technologies

Section 1.2 noted that the software part of holons must necessarily be an agent. Multi-agent systems in manufacturing are not an entirely unproven technology. They have been used in the past in systems where "data, control, expertise, or resources are distributed" or when legacy systems need to be inter-compatible [263]. Parunak explains that a primary advantage of agents is that not only can they decide locally as subroutines do but they also initiate their own (proactive) activity – thus enabling the holon properties [186]. In this way, agents give a framework for handling uncertainty, inconsistency, security and optimization [262], [261]. Parunak continues by stating that agents much like manufacturing entities have well-defined state variables that are distinct from their environment [188]. Hence, it is logical to make agents correspond to machines, tools, fixtures, products, parts, features and operations
More about agents can be found in Section 3.1.1.6 as well as on the FIPA website [87].

The intelligence of agents and the resultant agility on the manufacturing system creates new demands on the machine and device control in the shop floor. Specifically, these demands required lower level controllers to have real-time capability [16].

**Definition 2.3.1.** Real Time System – A system where its correctness depends not only on the logical result of computation but also on the time at which the results were produced [229].

Real-time systems (in order to exist) provide not just the necessary speed but also performance predictability [17]. The execution of tasks such as robot movements can be started and finished with temporal certainty. Real-time systems also implicitly demand reliability from their components to achieve this level of predictability [16]. They can also provide the necessary adaptability by sensing and responding to unexpected events [17].

The communication demands and responsiveness across a very distributed and decentralized architecture such as a holonic manufacturing system will require integral use of an internet backbone [269]. The method of utilization may vary from system to system but it is fundamental. For example, mobile agents are one innovative and effective use of the internet in HMS [90].

3 Developments in Holonic Manufacturing Systems

Having introduced holonic systems in Section 1, Section 2 gave an extended rationale for their adoption. Specifically, it showed how holonic systems aim to fulfill the faults of other manufacturing systems. It also rationalized holonic systems’ requirements and outlined general directions of technology development and implementation. This section builds upon this foundation with a review and elaboration of these development paths found in the literature. Throughout the course of the review, care is taken to evaluate the degree to which the holonic rationale and its requirements are fulfilled. The survey begins with the most prominent of the proposed system architectures in increasing degree of implementation. Next, methodologies for design and development of these architectures are discussed. Finally, the protocols, algorithms, and interactions upon which these architectures and methodologies are founded are investigated.

---

20Interestingly, much agent research has used functional agent decompositions as in [66], [167], [192], [63].
3.1 Architectures

The architecture section proceeds from a very general picture of holonic systems to one of greater detail. In other words, it investigates the most prominent conceptual descriptions and shows the manner in which realized implementations compare. Numerous architectures have been proposed as a result of the IMS feasibility program, and there is significant diversity, variation, and even disagreement in both conceptual descriptions and realized implementations. Additionally, a wide variety of architectures may be different at first glance but their differences may be resolved. To aid the review, the architectures must be differentiated into holon architectures that describe solely intra-holon components and HMS architectures that describe the component holons and their interactions. Laws notes that the literature is less than clear in making the distinction [141]. This review also attempts to highlight the repeatability, distinguishing features, and validation/evaluation processes (if any) of the mentioned architectures 21.

3.1.1 Conceptual Description

The first architectures published in the early and mid-1990’s published as a result of the IMS feasibility program were purely conceptual in nature. They improved upon the previously stated rationale solely with a description of a holonic manufacturing system’s component elements. In so doing, they formulate the system’s logical boundaries, clarify its requirements and give further rationale for its implementation.

3.1.1.1 Initial Conceptions

After the original conception of applying holonic principles to a manufacturing context [230], [231], and the creation of the HMS project [207], [112], Holonic Manufacturing Systems lacked a widely accepted concept of a holarchy structure and its individual holons. To solve this problem, Christensen proposed an initial holon architecture found in Figure 8 [62]. It may be viewed as the predecessor to the more detailed conception previously shown in Figure 1 on page 9 [61]. He also reiterated the holarchy structure found in Figure 7 on page 20 [61]. Finally, he attempted to standardize the nature of the holon interfaces that constitute a holarchy with Standards Theme Tasks (STT’s), as shown in Figure 9 [61]. At this point in HMS research, architectures were still very much conceptual but Christensen’s work did serve as the bedrock upon which other researchers’ works could coalesce.

21The validation/evaluation of an HMS architecture has been found to be non-trivial [264]
3.1.1.2 Cooperation Domains:
The cooperation domains paradigm was originally conceived soon after the beginning of
the HMS research program [72], [71], [73], [74]. Since then it has progressively developed and at times absorbed newly developing technologies [95], [91]. Nevertheless, its basic idea remains the same. Cooperation domains are where holons may location, contact, and interact with each other. In that sense, it may be viewed as a holarchy. Cooperation domains contain shared data structures and facilities for message passing. It also includes decision making and monitoring mechanisms that support the holon’s local activities. It also includes techniques and rules for treating compound holons. The research into cooperation domains solidified some of the more abstract descriptions before it and paved the way for more detailed architectures [95].

3.1.1.3 PROSA – Product, Resource, Order, Staff Architecture 

Soon afterwards, probably the most well-proliferated conceptual architecture was proposed. Called PROSA, it decomposed the shop floor into at least product, resource, and order holons as shown in Figure 10 [246]. The resource holon contains any physical resource required for manufacture at any level of production, i.e. raw material, robots, furnaces, AGV’s, a whole shop floor, or even an entire factory [245]. The product holon, holds the process\footnote{Process knowledge is information on how to perform a certain process on a specific resource [246].} and product knowledge\footnote{Product knowledge is information on how to product a certain product using a specific resource [246].} required for the proper manufacture of a given product [266]. In other words, it contains the end products of the traditional product design, process planning

![Figure 10: Basic Buildings Blocks of PROSA Architecture [246]](image-url)
and quality assurance departments [246]. The order holon represents a task in the holonic manufacturing system and manages the physical product proceeding through the system and maintains the process execution knowledge [267] 24. In addition to these holons, an optional staff holon may be added [32]. It provides the basic holons with supplementary information to make their decisions correctly [245]. It may be used to implement centralized algorithms deemed too difficult for a distributed solution [246]. PROSA also allows for specialization of the basic holons based upon their inherent characteristics, i.e. robot, CNC machine and conveyor resource holons [243]. The PROSA architecture also aggregates a large number of low-level holon in order to minimize its associate complexity [246]. This is done intuitively much like modern day plants aggregate machines into stations into cells, into shop floors. For example, Figure 11 shows how specialized product, resource and order holons may be embedded within a larger workstation resource holon.

Figure 11: Aggregation and Specialization in PROSA Workshop [245]

3.1.1.4 HCBA – Holonic Component Based Architecture:
HCBA, much like PROSA, can serve as a template for numerous implementations and applications of holonic systems [58]. It also uses a product and resource holon that are similar to those of PROSA [55]. The similarity, however ends there. HCBA grew from the application of Component Based Development (CBD) to HMS [56]. CBD is a software programming paradigm that is based upon plugging reusable and reconfigurable components together [233]. These components are given holonic attributes and go onto to form HCBA’s holons [58]. In

---

24Process execution knowledge is information on the current progress of a processes execution [246].
HCBA, the intra-holon structure is well specified from the bottom up and is shown in Figure 12 [55]. Each holon has components that correspond to machine, cell, and factory level components in conventionally controlled CIM systems [58]. Interestingly, this correlation is the foundation of a migration strategy to be introduced in section 3.2.3.1 [57]. Between the cell and machine levels, or in other words, at the hardware-software boundary, exists the BlackBoard System (BBS) [55]. The BBS is a cell-wide system that mirrors the real-time relays of the cell PLC and each holon is allocated a “slice” to control its respective hardware resource [58]. The BBS has the added advantage of providing a simulation interface for a virtual machine [55]. The message broker (MB) is responsible for inter-holon communication is one instantiation of the broker agents to be described in Section 3.1.1.7.1 [58]. The full holarchy of HCBA continues to use the CBD paradigm within a nested structure as seen in Figure 13 [55]. A manufacturing firm is decomposed into nested business, factory, cell and machine levels which double as the firms’ resource holons [58]. At each of these levels, also exists at least one product holon which dynamically spawns a WIP agent which coheres to
and escorts the product holon at the lower level [55]. For example, at the lowest level, a
subassembly creates a WIP agent for the machine level, which in turns escorts parts to and
through the machines for operation. This architecture was implemented at the University of
Cambridge’s Robot Assembly Cell and is further discussed in Section 3.1.3.2 [56].

3.1.1.5 HoMuCS – Holonic Multi-Cell Control System
HoMuCS is a generalize HMS architecture developed purely to create physically imple-
mented PROSA architectures [138]. This development process is described later the Ho-
MuCS methodology in Section 3.2.2.2. The functional models describe the generic holon
functionality as shown in Figure 14 [138]. In addition to autonomy and cooperation rules,
the functional models describe a holon’s basic functionality which includes dispatching, moni-
toring, planning, and auxiliary functions [139]. These functional models drive the functional
requirements for the object oriented models that form the architectures’s basic building
blocks [138]. From these models, product, resource and order holons are created [138]. The
product state models provide a template for centralized databases of product info [140]. The
HoMuCS architecture is explained further detail in [137] .

3.1.1.6 Agent Architecture Taxonomy
A vast number of multi-agent architectures exist in the literature and to review them would
be outside of the scope of this report. The objectives of the multi-agent system do not necessarily align with those of HMS [152], but it does nevertheless provide many of the software developments necessary for the implementation of HMS [39]. Hence, a basic agent and multi-agent system taxonomy with examples is provided for the reader of holonic manufacturing systems. A much more thorough review of multi-agent systems can be found in [216] and [217]. Additionally, attempts to standardize these agent and system architectures have been made [87][181].

3.1.1.6.1 Reactive Agents:
Maes defines single agent architectures as:

Definition 3.1.1. Agent Architecture – A framework that specifies how the agent can be decomposed into the construction of a set of component modules and how these modules should be made to interact. The total set of modules and their interactions has to provide an answer to the question of how sensor data and the current internal state of the agent determines the actions and future internal state of the agent [151].

Three basic types of agent architectures conform to this definition: reactive, deliberative and hybrid[141]. Reactive agents have a pre-specified action for every possible sensory input (from the environment or other agents), and if there exists more than one sensed input simultaneously, the actions or behaviours are completed in a pre-specified order of importance.
Brooks’ Subsumption Architecture is one such example [40], and is graphically represented in Figure 15. Intelligence is an emergent property of multi-reactive agent systems,

![Diagram of a Subsumption Agent: A Reactive Architecture](image)

but in general pro-active or goal directed behaviour is difficult to achieve [141].

### 3.1.1.6.2 Deliberative Agents

Deliberative agents may be viewed as the anti-thesis to reactive agents. They explicitly represent goals (or desires), formulate beliefs based upon accumulated sensory input, and develop plans (or intentions) to achieve those goals [43]. Rao’s Belief-Desire-Intention (BDI) architecture provides a good example of deliberative agents [197], and is diagrammatically shown in Figure 16. Deliberative agents can be both reactive and proactive but their implementation is impeded by the inability to decompose real-world events into an accurate symbolic representation [141]. Additionally, they may have difficulty responding in a timely fashion as they must both collect sensory input over time and compute through what may be a very complex logic [43].

### 3.1.1.6.3 Hybrid Agents

Hybrid architectures incorporate elements from both reactive and deliberative agents usually into a layered architecture [141]. Lower levels reactively respond to sensed inputs while higher levels formulate plans and cooperate with other agents [43]. One example is Muller’s InteRRaP architecture [177] and it is diagrammed in Figure 17. Hybrid agents are reactive, pro-active, autonomous and cooperate with other agents [141]. However, there remain difficulties in coordinating the inter-level interactions. Additionally, as a whole agent ar-
architectures lack a well formed design methodology [262]. Further examples of hybrid agent architectures may be found in [88].

3.1.1.7 Multi-Agent System Taxonomy:
There are many ways to group these agent architectures into a cohesive multi-agent system. These architectures may be divided into hierarchical, federation and autonomous agents.
approaches of which latter two are of greater interest in holonic manufacturing systems development [216]

3.1.1.7.1 Federation Architectures:
Federation architectures are further subdivided into facilitator, broker and mediator architectures but all have the defining feature of a central agent that coordinates the actions of the others [141]; much like a the staff agent in PROSA. For this reason, Weiss states that they may be implemented in HMS [258]. In the facilitator variation, facilitator agents receive incoming messages from a predefined group of agents and then translate, process and route these messages to the appropriate agent [211]. A diagram of the Desktop Utility Agent System is provided as an example in Figure 18 [65]. Broker architectures, as seen in Figure 19,

![Facilitator Architecture Diagram](image)

Figure 18: A Desktop Utility Agent System: A Facilitator Architecture [65]

are similar to facilitator agents but they also add monitoring and notification functionality and may be contacted by an local agent in the same system [211]. Mediators include all the

![Broker Architecture Diagram](image)

Figure 19: An Example Broker Architecture [211]

functionality of broker agents and extend their ability to coordinate the actions of more local
agents[141]. For example, they may use the recruiting mechanism as seen in Figure 20 to establish communication links between agents to later allow agents to communicate directly for the duration of their relationship [211]. Further examples federation architectures may be found in [21],[99], [166], [165], [164], [163], [162], [189], [219], [215]. The most prominent and perhaps most applicable of these is the metaMorph architecture developed at the University of Calgary [19], [17], [16], [18], [20], [36], [35], [168], [169], [214], [213], [210], [218].

Figure 20: An Example Mediator Architecture [211]

3.1.1.7.2 Autonomous Agent Architectures :
Autonomous agent architectures are different in that they do not rely on any centrally located agents [216]. In this sense, they are similar to the lowest negotiation levels of HCBA or PROSA (without staff holons). Shen defines them by their functionality. They must: 1.) not be controlled by another software agent or human being 2.) be able to communicate directly with and have knowledge about the environment and other agents 3.) have their own sets of goals and motivations [216]. Figure 21 gives a conceptual diagram [211].

3.1.1.8 Miscellaneous Conceptions :
A number of other conceptual architectures have been proposed to the HMS research community, but in general, they have yet to be deeply investigated [130], [178].

3.1.2 Detailed Design and Simulated Implementations
Section 3.1.1 provided many conceptual HMS architectures whose primary purpose was to delineate the logical boundaries of HMS components. This section describes holonic-type designs and simulated implementations. Specifically, they are based upon the full specification of an industrial problem into an HMS simulation. As they were developed at the same time
Figure 21: A Conceptualized Autonomous Agent Architecture [211]

as the conceptual descriptions, they often do not follow the same logical structures. Nevertheless, they do clarify some of the practical issues in HMS development and demonstrate some effective tools used in their resolution.

3.1.2.1 Rockwell/BHP Steel Rod Mill Water Cooling System:
The simulation of holonically controlled water cooled steel rod mill was one of the first tractable implementations of holonic principles. In a steel rod mill, hot steel coming out of the furnace is rolled, water cooled and then air cooled to a set temperature on a laying head as seen in Figure 22. The series of five water coolers is represented as separate negotiating holons [4]. The goal of the simulation was to demonstrate the system’s performance, robustness, autonomy, and flexibility [4]. Each cooling control holon was given functionality as illustrated in Figure 23 [170]. More physically, each cooling holon controlled a water cooling box, and a valve that restricts water inflow [170]. The system was simulated in a number of conditions: normal, one failed holon, one added holon, and modified temperature sensor [4]. With respect to the demonstration goals, the simulation showed that high performance was achieved in normal conditions and that robustness was maintained in partial failure conditions [4]. Additionally, the system demonstrated a high level of agility to new hardware configurations [4].

3.1.2.2 VTT Automation Robot Cell:
In 1997, VTT Automation designed a simulated holon manufacturing robot cell based upon
groundwork found in [115] and [113]. The distinguishing feature of this simulated architecture was the modularity and automation of their design implementation. The simulation was carried out within control and geometric environmental models; each residing on separate workstations connected by TCP/IP based LANs [114]. The geometric environmental model
residing on a Silicon Graphics Indigo Elan workstation uses the IGRIP 3D simulator software to create geometric models of the manufacturing devices [123]. It also uses the Graphical Simulation Language (GSL) code to manage the (virtual) machines’ state changes which are passed to the control model via TCP/IP sockets [114]. The control model is a graphics and text based and was developed using the Structured Analysis for Real-Time Systems (RT/SW) approach [257]. The CASE-tool SA-PROSA \(^{25}\) software package was used to create design graphics of composed of data flow and state transition diagrams [149]. These diagrams are later compiled directly into VHDL \(^{26}\) code using the VELVET compiler[182]. The VHDL code is then run on a Sun Sparc II workstation using the VHDL2000 simulator [194][114].

Further detail about the software development can be found at [136].

From a holonic manufacturing system point of view, the VTT simulation sits on a firm architectural base. As the work is in simulation, they specify the software components of the holon, but do take care to define the intra-holon machine interface as well as the inter-holon negotiation interfaces. They propose an agent is composed of proposer, planner, requestor, executor, monitor, and controller modules [113] which is an extension of the PEM concept proposed earlier in 1992 [115] \(^{27}\). A diagram of the basic function of this agent is shown in Figure 24 below. The interested reader is referred to the original text for further detail of the intra-agent interactions [114]. The agent architecture above is applied to each of the devices in the manufacturing robot cell: two Aitec ARS10 robots (with grippers, arms, and vision components), one injection molding machine, one welding machine, one conveyor, and two Automatic Guided Vehicles (AGV) [114]. The cell is graphically represented in Figure 25 [114]. Interestingly, Heikkila notes that the agents can be implemented in a hierarchical, heterarchical and multilayered structure, and ultimately the heterarchical structure is implemented [114]. A rough sketch of this structure is show in Figure 26 [114]. This simulation demonstrated controlled and resilient behavior in the presence of faults, and adapted to unpredictable situations effectively [114]. Heikkila does note however that network traffic grew rapidly with the number of holons (perhaps due to the heterarchical structure) [114]. This was corrected by redirecting messages to only a subset of holons deemed relevant a priori [114]. This may suggest the need for the aggregation or hierarchy found in the PROSA or

---

\(^{25}\)This software is not related to the PROSA Holonic Manufacturing System Architecture.

\(^{26}\)VHDL is the Very high speed integrated circuit Hardware Description Language [114]. A thorough text can be found at [144].

\(^{27}\)The PEM Model is composed of Planning Execution and Monitoring and are widely considered to be three of an agent’s primary tasks [115]. This is also consistent with the basic holon functionality in the HoMuCS architecture [138].
HCBA respectively.

3.1.2.3 Holden’s Engine Operations Engine Assembly Line The Australia based Holden’s Engine Operations (HEO) company implemented a simulated partial holonic manufacturing system for an engine manufacturing line using Simjava [176] and an Allen-Bradley Sll 5/03 PLC [101]. Their simulation requires a new control paradigm called Part-Oriented-

\[28\] HEO also decomposed the control problem in detail for a crankshaft machining state called the Huller-Hille machine [101]. However, the detailed design is neither simulated nor implemented.
Control [101]. In this paradigm, the manufacturing control centers around the evolution of a part from an initial to a final state through the use of services like translation and transfor-
mation. These services are then decomposed functionally into more basic services. In engine line, for example, the services are clamping, assembly, and transportation. These services are then associated statically or dynamically to their respective machine resources. Each instance of parts and resources is designated as a holon with message handling, self-diagnosis, and monitoring capabilities [101]. Additionally, part holons contain a specification of which series are required and how they are to be accomplished [101]. The resultant simulation demonstrated seamless PLC integration into Part-Oriented-Control, self-diagnosis capabilities and moderate messaging traffic [101]. As a result a (proprietary) prescriptive requirements specification was authored [101]. This detailed design is congruent with PROSA and HCBA and perhaps some of these prescriptive steps can be adapted to more sophisticated PROSA/HCBA designs.

3.1.2.4 Miscellaneous Designs A number of other virtual holonic manufacturing systems have been designed and implemented purely in software [179], [150], [75], [92], [93], [195], [226], [103], [161], [106],[107]. Many of these are mentioned later in Section 3.4 as their primary purpose was to illustrate innovative planning and scheduling algorithms.

3.1.3 Implementations
The architectures in the last section were only simulated. The following architectures have all been implemented on physical hardware. Many of them build off the conceptual description in Section 3.1.1 while others are entirely new like the simulated architectures in Section 3.1.2.

3.1.3.1 Daimler Chrysler Holomobiles In 2001, Daimler Chrysler fully implemented a physical holonic manufacturing system at its Mercedes Benz V6/V8 plan in Stuttgart, Germany [47]. The defining feature of this systems was the use of a number of holons to improve the performance of an existing conventionally controlled engine assembly line shown in Figure 27 [43]. Extensive analysis by simulation of the line showed that it was overly sensitive to machine and logical disturbances that caused deadlock and starvation [47]. In the event of elevated demand, it was not scalable to higher production level in the event of elevated demand [43]. Hence, the objective of the holonic system was to improve robustness and increase capacity in an incremental fashion [47]. As this was an industrial implementation, it was crucial that the implementation of the added holonic components minimize risk of investment and have a short time ramp-up to optimal performance [47]. Figure 28 shows the added holonic components. They were
implemented in a three step low risk migration path: 1.) flexible buffers are introduced to alleviate the effect of disturbances 2.) multi-function stations are added to give specific stations more operational capacity. 3.) Multi-function stations are converted to dedicated assembly stations and connected with transportation elements [43]. Each of piece of newly implemented hardware had its own appropriate software agent. The holonic components could be shut of at any time to leave a fully functional conventionally controlled line [47]. Analysis of the new system performance was done in simulation. The addition of holonic components resulted in 37% improved robustness and 61% improved throughput [47].

This holonic system clearly demonstrated the advantages of not just flexible equipment but also flexible software or agents to manage their decisions [43]. Most interestingly, the adaptability of the holonic components is used to effectively implement a low risk migration
This stated, this implementation more closely resembles a number of independent holons interacting with a decision-less conventional assembly line than a holarchy of holons negotiation amongst each other. These interactions may be flexible and temporary relationships but neither do they create a holarchy nor do they negotiate to find cooperative solutions.

3.1.3.2 University of Cambridge Robot Assembly Cell

The HCBA architecture was implemented at the University of Cambridge in a meter box robotic assembly cell [55]. The meter box comes in three color variations and is composed of three parts: A, B, and C, which from the box’s main base, frame and transparent plate respectively [57]. The full product is shown in Figure 29 and its diagram is shown immediately below it in Figure 30 [259]. The meter box is assembled by screwing part B on to its base part A [55]. Subassembly AB is then flipped so that part C can be screwed to it [57]. This is accomplished with the robotic assembly cell shown in Figure 31 [57] [30]. The PUMA robot picks and places the parts, so that the table can rotate them, and the Hirata can screws them together [57]. The flipper robot flips subassembly AB prior to screwing in part C [55].

Figure 29: A Holonically Assembled Meter Box [55][259]

---

29 Daimler Chrysler is putting greater emphasis in developing manufacturing systems with inherent agility. Another implemented purely agent-based manufacturing system called the Production2000+ can be found at [45][46].

30 This same robot assembly cell was used develop another very similar multi-agent system [126].
The holonic integration of HCBA is completed in static and dynamic phases which are shown in Figure 32 [55]. In the first, a resource controller, connected to an Omron PLC C200HE, forms the lower level of the holons’ software part [60]. It is responsible for the manufacturing execution as well as the house off the BBS [55]. In HCBA, resource holons may not negotiate directly with each other and hence the installation of the resource controller completes static integration [60]. Dynamic integration requires the installation of the production controller which is made up of the software component responsible for the negotiation and message brokering for both product and resource holons [55]. This results in the
fully implemented holarchy shown in Figure 33 and corresponds closely to the HCBA system architecture originally shown in Figure 13 on page 29 [60]. The product holon generates a WIP agent that escorts a physical part through the assembly cell. Further detail about this
implementation can be found in [53]. The primary advantages of this implementation is its separated systems integration sand its dynamic communication infrastructure which resulted in a functional holonic system [55].

3.1.3.3 KU Leuven  In 1994, KU Leuven implemented one of the first HMS prototypes [240]. In the years that followed, the flexible assembly system was used to demonstrate and complete a voluminous body of research. Among the primary results, was the previously mentioned PROSA architecture [246]. This testbed is diagrammatically shown in Figure 34 [241]. It consists of one PUMA 500 Robot (from Unimation), two Scara 7545 robots (from IBM), one Pragma A300 robot (from DEA) and one external transportation loops which together form the lowest level resource holons [241]. The system simultaneously assembles both electric switches and LEGO constructions using a PROSA type architecture [241]. A

Figure 34: Holonic Flexible Assembly System at KU Leuven [241]
centralized heuristics based reactive schedule is implemented with a staff holon and has been investigated throughway in [30], [31], [27], [26], [25]. The flexibly assembly system was also used to benchmark preliminarily holonic systems against hierarchical and heterarchical ones [242]. Many of their results, experimentally supported the arguments provided in Section 2. A number of fundamental research works on the implemented PROSA architecture can be found at [239], [24], [264].

3.1.3.4 Miscellaneous Implementations A number of quasi-holonic manufacturing systems that use embedded multi-agent systems have been reported. Some of these may be found at [6], [7], [51], [50], [98], [128], [147], [148].

3.2 Methodologies & Strategies

The methodologies section recognizes that a logical and step-wise framework is required for design, implementation and migration of holonic manufacturing systems. Many of the architectures presented in the previous section either were too conceptual to be very useful in a physical implementation or did not specify a detailed method by which their implementation could be repeated (perhaps with variations and/or in a different setting). Unfortunately, no fully comprehensive and applicable methodology has been written. This section, however captures some of the most useful tools for design and implementation of HMS. They are categorized as modelling tools, and design and migration methodologies and strategies.

3.2.1 Design & Modelling Tools

Holonic Manufacturing Systems research and development draws expertise from a wide variety of fields. Each of these brings its respective design and modelling tools to HMS development. This section highlights IDEF0 and Petri-Nets from industrial engineering as well as UML from object-oriented software engineering.

3.2.1.1 IDEF0:
IDEF0 or Integrated Computer Aided Manufacturing Definition is probably the most widely used method for modelling manufacturing system functionality [78]. It is composed of basic building blocks that illustrate a manufacturing activity’s inputs, controls, mechanisms, and outputs. A basic building block in addition to an illustrative example is shown in Figure 35 [108]. It is important to note that the links between IDEF0 blocks do not imply a flow of
information or material but rather imply a simple dependence [43]. Bussmann notes that IDEF0 networks have been used (with Petri-Nets) to create PLC programs [43]. IDEF0 modelling has been used in HMS development most notably in the HoMuCS architecture's functional models [265],[137]. Bussmann has also noted that it may be used as the input to the DACS multi-agent system design methodology [43].

3.2.1.2 Petri-Nets:
Petri-Nets is another widely used modelling technique. It is composed of places that represent conditions [199]. They are filled by tokens that denote the condition is true [199]. Finally, when the conditions become true they pass through a transition which leads to other places or states that may in and of themselves be conditions for further transitions [199]. An illustrative example shown in Figure 36 describe that a part, machine, robot need to be available for the robot to load the machine with part [199]. Petri-nets are particularly useful because they have a large body of theory that can be used to validate network properties [43]. Like IDEF0, Bussmann states that Petri-Nets can be used as the input to the DACS MAS design methodology [43]. Also, they have been applied at the KU Leuven and University of Cambridge laboratories [247], [59].

3.2.1.3 The Unified Modelling Language – UML:
Object-oriented programming has served an integral part in the development of holonic manufacturing systems partially because they share many of their principles in common [159]. For that reason, UML may be a useful HMS modelling language as it is an effective and simply way of capturing the salient features of an object-oriented software design [105]. While the language is fairly comprehensive of object-oriented language functionality, only
an illustrative class diagram used in PROSA is shown in Figure 37 [246]. A number of tutorials are available on the internet [105]. Classes such as the holonic manufacturing
system, product, resource and order holons are represented by boxes connected by three types of lines. Lines with diamonds represents that one class is contained within the other. Solid lines as those between product, resource and order holons show that they are associated with each other. Finally, lines with arrows (not shown) represent that one class is a specialization of another. One can imagine, for example a CNC robot holon connected to the resource holon block via an arrowed-line. Finally, the diagram is filled with number that represent the number of classes in that relationship. The notation 1* means ”at least one” [246].

3.2.1.4 Agent Building Tools :
Agent functionality goes well beyond the original intention of object-oriented languages. The functionality for agent implementation still exists, however, the software engineering process is less than straightforward. A number of agent building environments have been developed to facilitate the process. These include the open source JADE [125], [22], JACK by Agent Oriented Software Group [5],[126], and Zeus by BT [102].

3.2.2 Design Methodologies & Strategies
The review of HMS design methodologies centers around the following: ”There are major issues to be resolved including how to define holons for a given problem, what should be an appropriate holonic architecture, how to design effective cooperation mechanisms for good system performance, and implementation issues such as plug-and-play capability for fast deployment”. [104] Numerous partial attempts at an HMS design methodology have been made. Even Axiomatic approaches have been investigated [180]. However, all fall short in providing a prescriptive and easily usable comprehensive design methodology. This section, reviews some of these partial methodologies and strategies including those that may come from multi-agents systems research.

3.2.2.1 PROSA :
The KU Leuven research group in addition to the PROSA architecture provided a loose four step guideline for designing a holonic system [248]. In the first step, holons are identified. In the second, the system is designed in detail and implemented. Special care is taken to reuse system components wherever possible and to proceed in a bottom up approach according to the guidelines found in [248]. The third step provides for the installation and configuration of the HMS and finally the HMS is operate in the last step [32]. While these steps might seem intuitive, only the first is describe within the details of the PROSA architecture [248].
The identification of holons, however, is completed in a logical step-wise fashion using UML as seen previously in Figure 37 [246].

3.2.2.2 HoMuCS Methodology:

The HoMuCS methodology uses the HoMUCS architecture to create physically implemented architectures as seen in Figure 38 [138]. The methodology begins with an analysis phase which relies on the functional models to guide the basic holon design [137]. Additionally, the designer investigates the details of the functionality as dictated by his hardware specific requirements [138]. This process is basically one of extending the detail of the already existing functional models. In its original version, this process is completed using IDEF0 [137]. The detailed functional models dictate the detailed design of the object-oriented models [138]. These models can be described in UML as the HoMuCS methodology suggests [137]. Once again this is a process of instantiating and specializing the general holon classes provided in the HoMuCS architecture [138]. The product state models are then filled with product specific information [137]. Finally, these three models are then tied together in an integrated implementation. The HoMuCS toolbox facilitates this process with simulation and analysis tools [138]. This methodology is explained in full detail in [137] and is applied to a number of industrial case studies [206], [225].

Figure 38: An Overview of the HoMuCS Methodology [138]
3.2.2.3 DACS Multi-Agent System Design Methodology

The recently published Designing Agent Based Control Systems (DACS) design methodology for production control takes a very different approach from the modelling tools and methodologies that came before it. Its primary advantages are its great degree of usability by an industrial control engineer, its reuse of existing protocols, and its focus on a prescriptive approach based upon decision oriented models [43]. That said, it is a purely a MAS methodology, and any HMS features from its application is purely the result of the transferability of MAS concepts to HMS\textsuperscript{31}. The methodology follows three ordered and well-defined phases which begin after the specification of the plant’s mechanical components, arrangement and behaviour. In the first phase, the system’s control decisions are analyzed [43]. This is composed of identification of the systems’ decisions and then determining the dependencies between them [43]. In the second phase, the agents are identified [43]. This follows an iterative process: clustering the decision tasks into agents, and improving the decision model, if necessary [43]. Finally, the agent interaction protocols are chosen [43]. This consists of classifying the dependencies between agents, matching them to protocols in an existing protocol library, and customizing them for their use if necessary [43]. This methodology is explained in full detail and evaluated in [86].

3.2.3 Migration Methodologies & Strategies

The issue of migration from a conventional to a holonic system is critical [101]. There are comparatively few factories (green-field sites) being built across the globe and hence if holonic manufacturing sytems are to be widely adopted industrially, it must show a clear low-risk path to quickly upgrade a conventional system to a holonic one [48]. McFarlane describes a general migration framework in Figure 39 [172]. Figure 39a shows a conventional CIM hierarchy. Figure 39b depicts many of the HMS developments to date. They give distributed holonic solutions for a given level most likely without addressing other levels. The methods and strategies described in the following sections describe the later states in the framework. First, the planning, scheduling and execution layers and the control and physical plant layers are integrated separately into the software and physical parts of a holon. Eventually, the two parts are merged to form a full holonic solution [57].

\textsuperscript{31}It is the opinion of the author that Holonic Manufacturing System represent a specialized class of Multi-Agent Systems. Hence, a methodology that is sufficiently general for a Multi-Agent Systems will not be sufficiently prescriptive to describe a Holonic Manufacturing System.
3.2.3.1 HCBA Migration Strategy

The HCBA architectures were proposed with an inherent strategy to migrate from Figure 39b to the full holonic solution in 39 [54]. Its two-phase integration process is essentially a bottom-up migration strategy [57]. Holon formation starts at the machine level and crystallizes upwards to incorporate the standard CIM levels one by one [54]. At the incorporation of the planning and scheduling levels, the CIM realization of planning and scheduling are abandoned in favor of the product controller which provides the same functionality for each holon [55]. These steps are described in greater detail in [53].

3.2.3.2 Daimler-Chrysler Holomobiles Migration Strategy

The Holomobiles demonstration mentioned previously in Section 3.1.3.1 used a three phase implementation method [47]. Its migration path is inherently different from HCBA however because it assumes that the plant capacity needs to be augmented [47]. The augmented capacity can be implemented using HMS "plug-n-play" nature without detrimentally affecting the underlying conventional assembly line [43]. This migration, however, is not prescriptive enough to describe how it may be repeated. The implementation is also in a sense a hybrid
system because it has not implemented inter-holon negotiation. Hence, the migration path is not complete because it does not describe the full construction.

### 3.2.3.3 Holden’s Engine Operations Migration Strategy
Holden’s Engine Operations HMS simulation also provided a useful three-phase migration strategy [101]. In phase one, the services in the previously mentioned Part-Oriented-Control is decomposed [101]. Its feasibility is then determined in simulation by investigating 1.) the holon structure and boundaries 2.) the intra/inter holon messaging requirements, and 3.) the degree to which existing controllers can incorporate a Part-Oriented-Control/holonic framework [101]. In phase 2, the newly defined interfaces are implemented but no new functionality is added [101]. Finally, in phase 3, diagnostic, reconfiguration, and learning capabilities are inserted within the previously defined holon boundaries [101]. This seems like a feasible migration strategy that might be an important part of a more comprehensive methodology.

### 3.3 Protocols
Thus far, the developments section has given rather broad system wide views of holonic manufacturing systems. Section 3.1 identified the building blocks that form holonic manufacturing systems’ proposed architectures while Section 3.2 attempted to identify the methodologies by which to develop these architectures. Attention now takes on a finer scope and turns to the mechanisms that give the holons their desired behaviour. These are comprised of the protocols by which holons communicate amongst each other, the algorithms from which inter-holon behaviours emerge and the interactions by which holons drive their respective devices.

When speaking of protocols, one must differentiate between them and algorithms [114]. Protocols simply specify the communication method which maps state history to actions [203]. Strategies or algorithms are a particular method of using the protocol to achieve a particular result [203]. In this section, two particularly useful protocols are used.

#### 3.3.1 Contract-Net Protocol
The Contract-Net Protocol, originally proposed in 1980 is probably the most versatile and best-proliferated protocol in holonic manufacturing systems and multi-agent systems. It is a simple negotiation protocol that involves a manager and contractor [224]. The manager has
a task that requires completion [224]. He announces it to a number of contractors which bid on the task within a specified time frame [224]. The manager evaluates the bids, chooses the best one and awards the task to the contractor with the best bid [224]. Bussmann provides a simple graphical representation of the protocol and it is shown in Figure 40 below [43]. There

![Figure 40: The Contract-Net Protocol](image)

also exist numerous versions and extensions of the protocol such as the Extended Contract Net Protocol [89], the Market-Driven Contract Net [10], and the Levelled Commitment Contract Protocol [204]. One may also finds applications in many holonic manufacturing systems and multi-agent system algorithms [216] of which many are reviewed in Section 3.4.

### 3.3.1.1 The Tri-Base Acquaintance Model

The Tri-Base Acquaintance (3bA) Model may be viewed as a protocol or method to implement autonomous agents or holons at a minimum of network communication traffic [191]. One method of communication, (exemplified by a vanilla Contract-Net Protocol), is a general broadcast: one holon sends a message to all the other holons in the system [191]. This may be viewed as a completely connected web of nodes and the number of connections grows quadratically. Another method, typified by federation multi-agent systems requires the use of a central facilitator/broker/mediator [191]. This may be viewed as a star-shaped network and the number of connections grows linearly with the number of holons. The final method, the basis for the 3bA model, uses a facilitator-less acquaintance model where each holon maintains a certain level of social knowledge about other holons so as to minimize the number of information requests [191]. This proposed acquaintance model continues to use the Contract-Net Protocol but adds a three based knowledge structure [191]. The cooperation base maintains rather static infor-
mation such as the addresses and predefined responsibilities of the other holons in the system [191]. The task base includes information about what type of tasks a particular holon can complete a task as well as plans on how to carry them out [191]. This information changes a bit more frequently. The state base is in constant flux and maintains the load and state of other holons [191]. The intuitive idea of only requesting information via the network when it is needed forms the basis of the 3bA model. Implementation on a multi-agent system showed significant network traffic reduction [191]. Further detail about the multi-agent system implementation can be found at [153], [155], [154], [156], [190].

3.4 Algorithms

Having mentioned two of the more common protocols, their application in various algorithms can be explored. The algorithms surveyed in this section represent the temporary and flexible relationships in a holarchy’s inter-holon communication. From a CIM point of view, these algorithms fulfill the planning and scheduling functionality of CIM’s top two layers. Multi-agent planning and scheduling solutions have been proposed in recent years. Some integrate these two activities, while others separate them. For clarity, the distinction between the two activities is defined below.

Definition 3.4.1. Planning – The process of selecting and sequencing activities such that they achieve one or more goals and satisfy a set of domain constraints [216].

Definition 3.4.2. Scheduling – The process of selecting among alternative plans and assigning resources and times to the set of activities in the plan [216].

These algorithms distinguish themselves from conventional scheduling and planning algorithms by the requirements they must fulfill to exist within a multi-agent solution. McFarlane argues that they must have 1.) local (near) real-time reasoning capability, 2.) cooperative mechanisms between modular units, 3.) embedded models for autonomy 4.) diagnosis and error recovery routines and 5.) online learning capabilities [171]. Additionally, Bakers requires 1.) a balanced distribution of computational load among the agents, 2.) a physical correspondence between agents and their physical entities 3.) scalable growth of computational load versus physical entities [11]. The remainder of this section reviews the algorithms that best achieve these objectives. Although, integrated planning-scheduling algorithms better fulfill the criteria above, separated scheduling and planning algorithms are first treated as they can often be integrated later.
3.4.1 Scheduling Algorithms

Scheduling is one of the firm’s most important operations activities. It is also one of the most difficult and as a result lead times and works in progress levels are inflated, machines are under-utilized and production completion dates can not be controlled or predicted [150]. Industrially, MRP and MRPII packages are often used but as effectively planning software, they are ill equipped for detailed scheduling [172], [119]. Furthermore, they can not respond easily to on-the-shop-floor events and disturbances [119]. For detailed scheduling, heuristics such as critical ratio (CR) or shortest processing time (SPT) are often used instead [150][32]. However, they often yield suboptimal results whose performance is rarely measurable [119].

The crux of the problem is two fold. First, many scheduling problems are excessively computationally intensive; NP-complete in many instances [11]. (This might explain the industrial use of heuristics which at least guarantees a result in feasible time scale [11].) Secondly, even if a good schedule were found in a reasonable time scale, it would become invalid almost immediately as new jobs, rush orders and disturbances emerged [150]. Clearly, the algorithmic requirements previously mentioned would alleviate some of these difficulties. Within the scope of scheduling, the focus turned from finding computationally infeasible optima to finding measurable, consistent and near-optimal schedules within a reasonable time scale. A history of intelligent scheduling may be found in [193], while an excellent review of multi-agent scheduling algorithms can be found in [11]. Some of the most promising of these are reviewed here.

3.4.1.1 Lagrangian Relaxation Scheduling : 

Lagrangian relaxation is one such near-optimal advance scheduling algorithm [150][33]. It iteratively minimizes a job-tardiness objective function that can be easily decomposed into smaller problems; each corresponding to a shop floor resource or holon [119]. One author states that up to 90% of the algorithm is distributable[120]. It also gives a lower bound on the objective function which in turn can be used to measure the algorithm’s performance [150]. Additionally, its complexity is linear in both the number of machines and the length of the time horizon [119]. Hence, it represents a fully scalable solution. The algorithm is also reconfigurable as previous solutions can be used to seed new iterations that incorporate

32Many other heuristics such as Earliest Starting Time (EST) and Fewest Operations Eemaining (FOR) exist [117]. Heuristic scheduling methods have also been applied within a holonic context [117],[220],[232]

33Here the advance scheduling means that a number of tasks must be aggregated and then ordered and positioned in time in advance [11]. New tasks may be added on later iterations of the algorithm [119].
new orders, events and disturbances [120]. The algorithm has been used in numerous real life scenarios, including job-shops, and has shown to provide results within 10% of optimal [150]. One may find numerous examples of its application in [118], [120], [119], [68], [251], [146], [145], [52], [85] as well as later in Section 3.4.2.

3.4.1.2 Temperature Equilibrium Scheduling

Temperature equilibrium scheduling algorithms are an entirely different type of algorithm. It uses the concepts of relative heat, latent heat and temperature to equilibrate the work load between holons [75]. Each holon has a temperature and relative heat to the other holons [92]. "Hot" holons transfer some of their workload to "cooler" holons subject to a Contract-Net based protocol [75]. The transfer of tasks must also take into account the latent heat which may be defined as the price paid to transfer a task from one holon to another [75]. This is a near real-time (as opposed to an advance) scheduling algorithm [93]. It has been used in simulation to improve throughput, machine utilization and mean delay over more rule-based methods [92], [93].

3.4.1.3 Market-Driven Scheduling

Market-driven algorithm are probably the most common form of dynamic multi-agent scheduling algorithm. Numerous examples in multi-agent systems and holonic manufacturing systems have been applied in simulation but they are all variations on a theme [3], [11], [10], [12], [14], [13], [76], [121], [212]. All assume that the product order has already been decomposed into component processing tasks (planning)[11]. The principle feature is then to use the Contract-Net Protocol to correlate machines and tasks as new order come online in time [195] 34. Saad investigated one possible variation where resource holons contract product holons instead of vice versa [201]. Ramos and Sousa have also investigated market driven algorithms where the product holon is the manager [226]. They use a combined "forward influencing"/"backward influencing" method so that during bid calculation, each resources can evaluate both its availability and the effect of deadlines [227], [221]. This is analogous to the forward/backward scheduling algorithms found in industrial ERP/MES systems [11]. Conflict avoidance is also carefully treated [228]. Many other instances of market-driven scheduling exist and some of those that integrated planning functionality are treated in Section 3.4.2.2.

34The negotiation can also be done with strategies that lie anywhere from purely selfish or purely cooperative [114]. Examples of the can be found in [157] and [114] respectively.
3.4.2 Integrated Planning Scheduling Algorithms

In all cases, scheduling algorithms require as an input the parts’ component raw material and/or subassemblies as well as the type of machine/resources required to complete the processes that advance its production. This specification is by definition completed during the planning phase and can be completed in varying degrees of automation. Early in the life of CIM systems, bills of material (BOM) were recognized to greatly facility planning activities [108]. Later, CAD/CAM integration made its contributions [199]. However, there still remains a large gap between CAD/CAM outputs and planning specifications that may be used within a near real-time multi-agent system environment. Some authors have reported multi-agent based decomposition techniques [9],[23] while more recently the use of well-specified ”recipes” has been proposed to bridge the gap [173]. Additionally, many others have postulated reference architectures that envision this functionality [256]. The remainder of multi-agent planning developments have mostly reported as integrated planning-scheduling algorithms; the most prominent of which are reviewed below. It is useful to note that in most cases the planning functionality of one integrated algorithm can be implemented with different type of scheduling algorithm

3.4.2.1 Lagrangian Relaxation with Planning :

Lagrangian Relaxation scheduling algorithms can acquire planning functionality with a distributed two-phase ”constraint satisfaction” planning method based upon the contract-net protocol and an adaptive consistence algorithm developed in [70] [103]. In phase one, part-assembly-part triples are generated. These triplets are then combined by satisfying contact constraints using a contract-net protocol to form a ”globally consistent combination” or a fully assembled product. Phase two takes this combination and parses it into an assembly sequence and identifies the machine types that can perform each assembly operation. The output of the planning can then be used normally in Lagrangian Relaxation [103]. This combined algorithm has been implemented on a holonic architecture originally described in [110] using a simulated testbed composed of three different assembly operations, one AGV, and products. It was show to generate efficient plans and schedules [103]. A similar demonstration based upon a Toshiba facility has also been demonstrated [104].

35In this demonstration, planning refers to matching appropriate parts and assembly plans to products [103].
3.4.2.2 Market Based Approaches

Market based integrated planning/scheduling approaches differ from pure market driven schedules by the addition of some decomposition method which is specified to varying degrees of detail in the literature [131], [249]. In one simulated implementation, order are accepted based upon a management agent’s internal model of the expected profit for a given order [157]. The same agent is able to directly translate the order into a sequence of tasks which are then negotiated upon by a machine resources [157]. In another simulated implementation, design agents negotiate with manufacturting agents to fulfill features’ specifications and tolerances [161]. Upon determination of feasibility, the design agent provides a list of (machine-type specific) tasks to a "matchmaker" agent which assigns them to machine agents via a contract net protocol [161]. A cost model based upon machining, tool change, and setup times is used to determine the value of the bids [161]. Yet another simulated implementation is done almost identically but with the usage of a "STEP" which includes product details including geometry, dimensions and functionality. A "STEP" parser and interpreter are then used to create a task file to be used in the planning process [106],[107].

3.5 Interactions

The previous section described the algorithm that managed inter-holon communication. This section introduces some of the more promising interactions necessary for intra-holon communication.

3.5.1 IEC 61499 Function Blocks

Implementing real-time functionality in a holonic manufacturing system is difficult because it is rarely straightforward to reconfigure part flows and interactions while maintaining real time constraints [90]. The IEC 61499 function block standard provides one proposed solution [234]. It builds upon the IEC 61331-3 standard for PLC and provides an encapsulated object with execution and algorithmic capabilities. Its basic structure is divided into algorithm execution and execution control halves which are shown in Figure 41 [39]. The former functions on a data flow of sensed events similar to other function block models [39]. The latter uses an event flow to trigger the algorithm execution [39]. Brennan captures the functionality in the following sequence and its associated diagram in Figure 42 [39].

---

36The standard PLC scan cycle creates significant inflexibility in executing scheduled events in a real-time fashion [39].
1. Relevant input variables values are made available.

2. The event at the event input occurs.

3. The execution control function notifies the resource scheduling function to schedule an algorithm for execution.

4. Algorithm execution begins.

5. The algorithm completes the establishment of values for the output variables.

6. The resource scheduling function is notified that algorithm execution has ended.

7. The scheduling function invokes the execution control function.

8. The execution control function signal an event output [39].

The advantages of IEC 61499 function blocks are many. Besides integral real-time functionality, their modularity lends themselves to distributed control systems [96]. They may also be aggregated in numerous flexible combinations [96]. Their encapsulation also allows for modular development (hence lower costs) of all holonic systems activities (i.e. sensing, actuation, control, communication) [96]. System integration costs will also be reduced as compatible software will be integrated; potentially with open-standard methodologies [96]. From a migration viewpoint, they provide an open-standard modelling framework which
can be used, for example, to reimplement academic prototypes to proprietary systems [96]. Hence, IEC 61499 function blocks can be viewed as a good way to implement agents, holons, and their respective systems [96]. As the IEC 61499 has developed, it has received a great deal of attention within the MAS and HMS community. There have been many investigations into the use of IEC 61499 Function blocks [37], [253], [254], [252], [255], [272], [271], [275]. In particular, improvements in fault tolerance [94], and reconfigurability have been investigated [273], [274].

3.5.2 Function Block Operating System

The use and implementation of IEC 61499 function blocks is facilitated with the creation of a function block operating system (FBOS) [96]. Fletcher identifies two paths for achieving holonic reconfigurability [96]. In the first, the function blocks are coupled tightly to an object-oriented language [96]. This uses conventional computing models and operating systems to give real-time functionality [96]. For example, one could simulate the function block in JAVA, create a UML model with real-time and agent extensions prior to the eventual function block implementation [96]. The other option is the use of a FBOS to schedule and guarantee the timing of function block execution [96]. A FBOS would also share the manufacturing resources among conceivably numerous function blocks [96]. It can also encapsulate the function blocks to create a virtual machine that simulates the entire holonic manufacturing system [96]. One possible vision of a FBOS based HMS architecture is found in Figure 43. Finally, the FBOS can include a Real Time Guarantee System (RTGS) that directs interfaces
with manufacturing resources in order to enforce execution time periods and deadlines [96].

4 Open Issues & Industrial Adoption

After having reviewed the architectures, methodologies, protocols, algorithms and protocols in the Holonic Manufacturing System field, one can see a number of research gaps remain that prevent full industrial adoption. They can be divided grossly into design and implementation gaps.

4.1 HMS System Design

4.1.1 Availability of Proven Design Methodologies

As seen in the methodologies section, the holonic manufacturing systems field lacks a truly prescriptive, easily implementable design methodology. The DACS methodology as a multi-agent system methodology provides a good reference point but can not be taken in its present
form to design holonic manufacturing systems. Additionally, the DACS methodology, HCBA architecture, and the Daimler-Chrysler Holomobiles demonstration together provide the beginnings of a migration design methodology for an existing CIM architecture. Finally, the Holonic Manufacturing System Consortium will undoubtedly have to investigate in detail the existing CIM design methodologies as a litmus test for its own methodologies. Any enterprize when given the option of a developing manufacturing system will only choose an HMS design methodology if it has been favorably evaluated against a proven CIM methodology.

4.1.2 Analysis of Performance of Holonic Manufacturing Systems

This leads to the second unresolved design point. The architectures above, theoretical or implemented need to be systematically compared and evaluated against the current best alternative. Without such evaluation rigor, industrial partners will perceive the investment risk as too high for full implementation. Specifically, the HMS community has often stated that a holonic system is robust and adaptable to disturbances and failures. These particular advantages should be explicitly demonstrated.

4.2 HMS System Implementation

4.2.1 Appropriate environments for implementing holonic control

As a result of the evaluation and analysis done in the previous section, there should be a set of upon which an HMS system will most successfully function. It is unlikely that the agility of holonic manufacturing systems is required for all industrial sectors. Hence, the evaluation studies should give guidelines as to when best implement holonic systems. This may include an identification of the most suitable industrial sectors, product types, and market behaviors.

4.2.2 Establishing suitable standards for holonic control systems

Finally, there must be a greater degree of standardization of holonic systems. As reviewed, there exists a very large number of holonic architectures among which lies a great deal of disagreement. Greater agreement on holonic architectures, perhaps aided by a solid evaluation process, must be achieved. Ultimately, there may be many different HMS architectures as there are many CIM architectures. However, the designer should have access to a library of standardized HMS components. Certainly, the FIPA and IEC 61499 function blocks standards provide a promising point of development in this regard.
References


