

Alpha-Q-Routing

部分シミュレーションと推定時間テーブルを組み合わせた経路探索
Route Search Combining Partial Simulation and Estimated Time Tables

遠藤彬史*¹ 村田昇*¹ 丸山祐丞*² Felix Beierle*²
Akifumi Endo Noboru Murata Yusuke Maruyama

*¹早稲田大学 Waseda University *²Eaglys 株式会社 EAGLYS Inc.

Efficient route control in automated material handling systems, particularly Overhead Hoist Transport (OHT) systems, is a critical challenge in semiconductor manufacturing facilities. As production line conditions constantly change and unexpected congestion occurs, conventional shortest path algorithms based on mathematical optimization struggle to effectively avoid traffic congestion. To address this challenge, this research proposes a route control method called "Alpha-Q-Routing." Unlike conventional Q-learning-based methods that only utilized edge-level partial simulation results for Q-value updates, this method combines partial simulation results for multiple candidate routes with estimated travel times from Q-tables, enabling route selection based on more accurate arrival time predictions. Numerical experiments in an environment with 50 vehicles demonstrated that the proposed method achieved the highest number of completed deliveries compared to conventional methods, particularly proving its superiority in environments with high vehicle density. This achievement contributes to realizing efficient logistics management in dynamically changing manufacturing environments.

1. はじめに

半導体製造工場 (fab) において、自動搬送システム、特にオーバーヘッドホイス搬送システム (OHT) は製造プロセスにおいて重要な役割を果たしている。近年の半導体需要の急増に伴い、fab の規模は拡大し、それに伴って必要な OHT 車両数も増加している。製造ラインの状況は常に変化し、予期せぬ混雑や機器の故障が発生するため、システムの動的な性質への対応が必要となっている。従来の数理最適化に基づく静的な最短経路探索アルゴリズムでは、渋滞状況を考慮した経路選択の計算負荷が高く、実環境に適用することが難しい状況である。また、fab の製造装置は日々更新・入れ替えが行われ、物流パターンが常に変化するため、従来のアルゴリズムではレイアウトや物流パターンの変更に対して柔軟な対応が困難であり、パラメータ等の調整が煩雑になる課題がある。これらの背景を踏まえると、OHT システムにおいて、動的な混雑状況に適應でき、かつ実用的な計算時間で実行可能な経路探索手法が必要である。本研究では、半導体製造工場の大規模 OHT システムにおいて、部分シミュレーションと Q テーブルを組み合わせることで実用的かつ効果的な経路制御を行うアルゴリズムを提案する。次に、シミュレーターを用いた実規模の検証実験の実施と、従来手法および先行研究との性能比較を通じて、その実用性を検証する。

2. システム構成とノードベースの制御

OHT システムは、OHT 車両、天井に設置された搬送用レール (OHT トラック)、および搬送システム上の決定ポイントであるノード (ステーションノード、分岐・合流ノード) で構成される。このアプローチでは、各ノードが意思決定エージェントとして機能し、到達した車両の目的地確認、経路評価、最適経路の選択、および移動指示を行う。これにより、大規模システムにおける分散的な意思決定が可能となり、システム全体

の計算負荷を軽減できる。

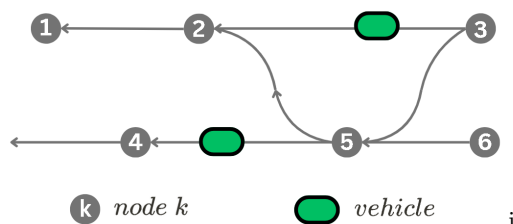


図 1: ノードベースの制御の概念図

3. 関連研究

3.1 数理最適化での解法

ダイクストラ法や A*アルゴリズムなどの数理最適化アプローチは理論的には優れた解を導出できるが、大規模な OHT システムでは計算時間が膨大になり、リアルタイムでの経路最適化が困難になることや、予期せぬ状況変化への迅速な対応が難しいなど実用上の問題がある [Hwang 19]。ヒューリスティックなルールを追加して、効率化を図ることも考えられる。しかしこれもシステム規模や車両数によって最適なパラメーターが変化することや、渋滞の定義 (エッジ上の車両数の閾値など) が工場規模に依存することによって、経験的な調整が必要となり、自動化が困難になってしまう。これらの課題は、実際の工場環境における経路選択システムの実用化を妨げる重要な要因となっている。

3.2 Q-Routing

先行研究の一つである Q Routing [Hong 22] は、強化学習の Q 学習を用いた動的経路制御アルゴリズムである。この手法では経路の厳密解を求めずに Q テーブルの値を用いて近似解を短時間で求める。設計する Q テーブルの Q 値 $Q[(d, i), j]$ は、現在のノード i から目的地ノード d までの推定走行時間を表し、次のノード j に移動する場合の値として定義される。

連絡先: 遠藤彬史, 早稲田大学, 東京都新宿区大久保 3-4-1, endoakifumi@akane.waseda.jp

つまり、Q 値は OHT 搬送機の後続経路における経路長と混雑レベルを考慮した推定移動時間を表す。テーブルは全ての目的地、現在地、選択肢の組み合わせで作られる。

destination (d)	current node (i)	next node (j)	Q Value [(d,i),j]
1	3	2	20
1	3	5	35

図 2: Q-Routing における Q テーブルの設計

エッジの走行が完了するたびに、現在のノード i から次のノード j までの実際の走行時間 $t(i, j)$ を報酬として、すべての目的地に対して異なる TD 誤差を作り、Q 値を更新する。報酬と TD 誤差の構成は以下の式で表される：

$$R[(d, i), j] = t(i, j) + \rho\phi(d, j) - \phi(d, i) \quad (1)$$

$$\begin{aligned} \delta[(d, i), j] &= R[(d, i), j] \\ &+ \min_k Q[(d, j), k] - Q[(d, i), j] \end{aligned} \quad (2)$$

ここで $\phi(d, i)$ はノード i から目的地 d までの静的最短走行時間を表し、 ρ は影響度を定めるパラメータである。

この手法では更新の効率を向上させるために、適格度トレースを変形して、マルチステップヒューリスティック適応法を提案している。ノード j からエッジ (l, m) までの静的最短距離 $\Phi(j, l, m)$ により、TD 誤差 $\delta[(d, i), j]$ を距離に応じて減衰させる。減衰係数 $\lambda_1 \in (0, 1)$ 、 $\lambda_2 > 0$ を用いて、遠方のエッジほど更新の影響が小さくなるように関連するエッジに $t(i, j)$ からの情報を伝播するように設計されている。Q 値の更新は以下の式で表される：

$$Q[(d, l), m] \leftarrow Q[(d, l), m] + \alpha\delta[(d, i), j]\lambda_1^{\Phi(j, l, m)/\lambda_2} \quad (3)$$

更新は以下のようなアルゴリズムとして表せる：

Algorithm 1 Multi-step Q(λ) updates

Require: エッジの $t(i, j)$

```

1: for 全ての目的地  $d$  do
2:    $R[(d, i), j] \leftarrow t(i, j) + \rho\{\phi(d, j) - \phi(d, i)\}$ 
3:    $\delta[(d, i), j] \leftarrow R[(d, i), j] + \min_k Q[(d, j), k] - Q[(d, i), j]$ 
4:   for  $(l, m) \in (i, j)$  に影響を受けたエッジ do
5:      $Q[(d, l), m] \leftarrow Q[(d, l), m] + \alpha\delta[(d, i), j]\lambda_1^{\Phi(j, l, m)/\lambda_2}$ 
6:   end for
7: end for

```

車両がノードに着いた際に、現在ノード i から選択可能な各次ノード j に対して、目的地 d までの推定所要時間 $Q[(d, i), j]$ を参照し、ボルツマンソフトマックス方策により、Q 値から選択確率を計算する。

$$\pi[j|(d, i)] = \frac{e^{-\frac{1}{\tau}(Q[(d, i), j] - \bar{Q}(d, i))}}{\sum_k e^{-\frac{1}{\tau}(Q[(d, i), k] - \bar{Q}(d, i))}} \quad (4)$$

ここで $\bar{Q}(d, i)$ は現在ノード i から選択可能な全行動の Q 値の平均である。温度パラメータ τ により探索と活用のバランスを調整できる。

3.3 Active-Q-Routing

Active-Q-Routing[Park 24] は、Q-Routing を拡張した動的経路制御アルゴリズムである。基本的な構造は Q-Routing と同じだが、車両がノードに到着した時点で、実際に通過する前に前方のエッジの走行の予測時間をつかって Q 値の更新を行うことで、周辺の渋滞状況を考慮した、より効率的な経路選択が可能となる。報酬の計算式は以下のように $t(i, j)$ が $\hat{t}(i, j)$ に置き換わる：

$$\hat{R}[(d, i), j] = \hat{t}(i, j) + \rho\phi(d, j) - \phi(d, i) \quad (5)$$

また、実際に通るかに関わらず分岐のエッジそれぞれに対しての予測により更新を行うため、片方の経路の Q 値が上がった際にその経路が選ばれず、渋滞情報が回復しない問題にも対処している。エッジ間の時間の予測では、予測対象となるエッジの周辺で、渋滞に影響を与える可能性のある車両のみを含む部分的なシミュレーションを実行して対象エッジの走行時間を予測する。

4. 提案手法 Alpha-Q-Routing

本稿では、近距離では部分シミュレーションによる詳細な渋滞評価を行い、遠距離では推定時間テーブルを用いるハイブリッドな経路選択手法 Alpha-Q-Routing を提案する。経路選択は以下の手順で実行される：

1. 一定距離内の経路パターン列挙
2. 静的評価による上位パターンの選択
3. 選択パターンへの部分シミュレーション実行
4. 最適経路の決定

本手法は、探索の深さ N とシミュレーション数 M という 2 つのパラメータによって制御可能である。パターン列挙では、現在地からノード数 N までに到達可能な全ての経路 $p^{(k)} = (i, j_1, \dots, j_n)$ を列挙し、その集合を P とする。静的評価では、各パターン $p^{(k)}$ について、他の車両が存在しない理想状態での移動時間 $\text{staticTime}(p^{(k)})$ と、終端ノード j_n から目的地 d までの最小 Q 値 $\min_k Q[(d, j_n), k]$ の和を評価値 $v^{(k)}$ として算出する。各パターンに対して、以下の評価値 $v^{(k)}$ を計算する：

$$v^{(k)} = \text{staticTime}(p^{(k)}) + \min_k Q[(d, j_n), k] \quad (6)$$

$v^{(k)}$ の上位 M 個のパターンを選択し、集合 S とする。この過程により、目的地への接近性が高く、不要な迂回を含まない有望なパターンを効率的に選別することが可能となる。

動的評価では、選別された集合 S の各パターンに対して部分シミュレーションを実行する。各パターン $p^{(k)}$ についてのシミュレーション時間を $\text{simTime}(p^{(k)})$ とする。これと $\min_k Q[(d, j_n), k]$ を用いて以下の評価値 $v'^{(k)}$ を計算する：

$$v'^{(k)} = \text{simTime}(p^{(k)}) + \min_k Q[(d, j_n), k] \quad (7)$$

最終的な経路選択として、動的評価値 $v'^{(k)}$ が最小となるパターン $p^{(k)}$ を採用する。

また、シミュレーション結果のうち $t(i, j_1)$ を Q 値の更新にも利用することで、将来の経路選択精度の向上を図っている。以下のアルゴリズムは目的地を含まない場合の処理である：

Algorithm 2 Alpha-Q-Routing ($d \notin$ 探索範囲)

Require: 現在地 i , 目的地 d , 探索の深さ N , シミュレーション数 M
 $P \leftarrow N$ ノード以内の全ての経路パターン $p^{(k)} = (i, j_1, \dots, j_n)$
 2: **for** $p^{(k)} \in P$ **do**
 $v^{(k)} \leftarrow \text{staticTime}(p^{(k)}) + \min_k Q[(d, j_n), k]$
 4: **end for**
 $S \leftarrow \text{TopM}(\{v^{(k)}\}, M)$
 6: **for** $p^{(k)} \in S$ **do**
 $v^{(k)} \leftarrow \text{simTime}(p^{(k)}) + \min_k Q[(d, j_n), k]$
 8: **if** $t(i, j_1)$ による更新がまだ行われていない場合 **then**
 $t(i, j_1)$ を用いて Algorithm 1 の更新を実行する
 10: **end if**
end for
 12: **return** $p^{(k)}$ that minimizes $v^{(k)}$

先行研究では多くのパラメータ調整が必要であったが、本手法では探索の深さとシミュレーション数という整数値での直感的な設定が可能である。探索距離は詳細評価を行う範囲を、シミュレーション数は計算コストと精度のトレードオフを表しており、システムの挙動が予測しやすく実環境への適用が容易である。

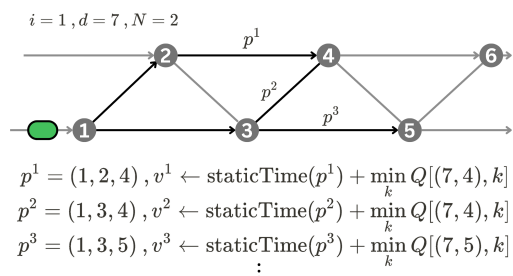


図 3: パターン列挙, 静的評価の概念図

5. 実験

5.1 実験設定

提案手法の有効性を評価するため、ダイクストラ法、Q-Routing、Active-Q-Routing、Alpha-Q-Routing の比較実験を実施した。実験は半導体製造工場における大規模 OHT システムを想定し、工場内の運搬経路をグラフ構造として表現した。シミュレーター環境は以下の通りである：

- システム構成
 - OHT 車両：材料を搬送する自動搬送車両
 - OHT トラック：天井に設置された搬送用レール
 - ノード：システム上の決定ポイント（ステーションノード、分岐・合流ノード）
- 動作制約
 - 衝突回避制約：合流部での車両同士の衝突防止
前方車両との安全距離確保
 - 車両容量制約：各車両は 1 度に 1 つのみ搬送可能

実験は 5000 ステップにわたり、平均輸送時間、総配達完了数、計算時間を評価指標として採用した。システムの規模による影響を評価するため、車両数を 10 台、25 台、50 台と変化させて検証を行った。実験の再現性確保のため、配送タスク生成の乱数シードは固定値を使用した。ダイクストラ法では、経路の再計算は行わず、静的な移動時間を元に算出した経路を用いている。

5.2 実験結果

実験結果を 1 に示す。

表 1: 実験結果の比較

指標	ダイクストラ法	Alpha-Q-Routing	Active-Q-Routing	Q-Routing
車両数: 10 台				
輸送時間	40.04	40.08	38.97	39.93
配達完了数	199	198	200	195
計算時間	0.28	4.65	6.10	2.04
車両数: 25 台				
輸送時間	45.16	45.59	45.26	46.11
配達完了数	449	456	446	434
計算時間	0.72	21.92	13.04	4.50
車両数: 50 台				
輸送時間	61.39	60.02	63.49	67.77
配達完了数	628	667	631	546
計算時間	1.45	78.69	21.80	7.46

5.3 考察

Active-Q-Routing は、学習を通じて渋滞の傾向を捉えることができ、前方エッジの予測により環境変化への適応速度が向上している。しかし、これらの手法には学習率や softmax 温度パラメータなどのハイパーパラメータ調整が必要であり、車両密度によって最適値が変化するという課題がある。50 台規模の環境では、過度な渋滞状況下で Q 値の推定が不安定になり、適切な経路選択が困難になる傾向が見られた。提案手法の Alpha-Q-Routing は、部分シミュレーションによる詳細な状況予測と、静的評価による大域的な経路選択を組み合わせることで、さらなる性能向上を実現した。特に 50 台規模の環境において、最も多い 667 件の配送を完了し、平均輸送時間も 60.02 秒と最短を記録した。これは、局所的な渋滞予測と大域的な経路計画を効果的に統合できた結果と考えられる。また、シミュレーションモデルの変更や拡張が容易であり、工場レイアウトの変更や新しい運用ルールの追加にも柔軟に対応できる点は、実環境への適用において重要な利点となる。しかしながら、提案手法では車両数の増加に伴って計算時間が著しく増大する傾向が確認された。これは部分シミュレーションのコストが車両数に大きく依存するためである。実際の工場環境での運用では、制御システムは実機との同期を維持しながら動作する必要があり、計算時間の短縮が重要な課題となる。この課題に対しては、各経路候補に対する部分シミュレーションを並列処理することで、大幅な計算効率の向上が期待できる。

6. まとめ

本研究では、半導体製造工場の大規模 OHT システムにおいて、部分シミュレーションと Q テーブルを組み合わせた新しい経路制御手法「Alpha-Q-Routing」を提案した。従来手法と比較して、特に高密度環境下での搬送効率が向上し、より多くの配送タスクを完了できることを実験で示した。また、探索距離とパターン数という 2 つの整数値パラメータによる直感的な制御が可能となり、従来手法で必要とされていた複雑なパラメータ調整の課題を解決した。

一方で、車両数の増加に伴う計算時間の増大が課題として明らかになった。今後は部分シミュレーションの並列処理化による改善や、より大規模な環境での検証を進めることで、実環

境での実用化を目指す。本研究の成果は、半導体製造工場における自動搬送システムの効率化に貢献し、製造プロセス全体の生産性向上につながることを期待される。

参考文献

- [Hong 22] Hong, S., Hwang, I., and Jang, Y. J.: Practical Q-Learning-Based Route-Guidance and Vehicle Assignment for OHT Systems in Semiconductor Fabs, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 35, No. 3, pp. 385-396 (2022)
- [Hwang 19] Hwang, I. and Jang, Y. J.: Q-learning-based dynamic route guidance algorithm for overhead hoist transport systems in semiconductor fabs, *International Journal of Production Research*, Vol. 58, No. 4, pp. 1199-1221 (2019)
- [Park 24] Park, J., Shin, D., Hong, S., Hwang, I., Hwang, S., Jang, Y. J., Lee, J., and Lee, J.: Autonomous Robot Orchestration Solution for OHT with Machine Learning and Digital Twin, *IEEE Advanced Semiconductor Manufacturing Conference* (2024)