Fundusze Europejskie dla Wielkopolski
Dofinansowane przez Unię Europejską
SAMORZĄD WOJEWÓDZTWA WIELKOPOLSKIEGO

Automatyka i robotyka

Skrypt dla studentów studiów niestacjonarnych

Rok III

redakcja

Robert Cieślak

Konin 2026

Tytuł

Automatyka i robotyka
Skrypt dla studentów studiów niestacjonarnych
Rok III

Redakcja naukowa

Robert Cieślak

Autorzy

Bronisław Bryl
Andrzej Milecki
Monika Muszyńska
Robert Roszak
Paweł Sobczak

Projekt pn.
„Rozwój studiów o profilu praktycznym i form kształcenia ustawicznego dostosowanych do potrzeb Wielkopolski Wschodniej”,
realizowany przez Akademię Nauk Stosowanych w Koninie,
jest współfinansowany przez Unię Europejską
ze środków Funduszu na rzecz Sprawiedliwiej Transformacji
w ramach programu Fundusze Europejskie dla Wielkopolski 2021-2027

Wydawca

Akademia Nauk Stosowanych w Koninie
ul. Przyjaźni 1, 62-510 Konin

Spis treści

Symulacja układów złożonych z uwzględnieniem zestawów kontaktowych w procesie modelowania i symulacji

dr hab. inż. Robert Roszak, prof. ANS

Modelowanie i symulacja z wykorzystaniem metod komputerowych to proces tworzenia cyfrowej reprezentacji (modeli) rzeczywistych układów i fizycznych zjawisk, tak aby następnie analizować, przewidywać lub optymalizować ich zachowanie za pomocą symulacji komputerowej. Modelowanie to tworzenie reprezentacji, a symulacja to analiza numeryczna tego modelu na komputerze. Celem jest wizualizacja, w jaki sposób zachowa się model w różnych warunkach. Obszar modelowania i symulacji możemy podzielić na:

- Proces modelowania - to proces tworzenia uproszczonej, często matematycznej i fizycznej, reprezentacji rzeczywistego systemu lub zjawiska. Modelowanie może obejmować tworzenie modeli matematycznych, graficznych, logicznych lub innych reprezentacji, które odzwierciedlają najważniejsze aspekty badanego zjawiska.
- Proces symulacji - to uruchamianie modelu na komputerze i obserwacja jego zachowania w różnych warunkach, często w celu przewidywania zachowania systemu lub optymalizacji jego działania. Symulacja może być używana do badania wpływu różnych zmiennych na system, testowania różnych scenariuszy i podejmowania decyzji.

Główne zastosowania modelowania i symulacji komputerowej to dziedziny nauki takie jak fizyka, inżynieria, biznes, medycyna, nauki społeczne i wiele innych. Do najważniejszych zalet stosowanie narzędzi modelowania i symulacji zalicza się:

- Przewidywanie zachowania systemów przed ich budową lub uruchomieniem,
- Optymalizację działania systemów.
- Testowanie różnych scenariuszy i podejmowanie decyzji.
- Ułatwienie zrozumienia złożonych systemów.
- Zmniejszenie kosztów i ryzyka związanych z prototypowaniem i testowaniem realnych systemów.

Jedną z powszechnie stosowaną metodą w procesie symulacji jest Metoda Elementów Skończonych (MES). MES ta jest jedną z najczęściej wykorzystywanych narzędzi w obliczeniach numerycznych polegające na podziale układu ciągłego na układ dyskretny opartego na tzw. elementach skończonych, wykorzystując lokalny układ współrzędnych oraz proste funkcje wielomianowe oraz macierz Jacobiego. Metoda ta pozwala transformować zdeformowane elementy skończone w proste i jednakowe formy upraszczając przy tym obliczenia. MES zastosowano początkowo w mechanice, ale bardzo szybko znalazła zastosowanie w innych dziedzinach, m. in. w analizie przepływu ciepła. Warto podkreślić, że system (program) oparty na MES oferuje wiele alternatywnych możliwości budowania modelu i równie wiele parametrów rozwiązania. Oferuje też narzędzia do badania wyników i testowania ich jakości. Jedynym przewodnikiem umożliwiającym dokonywanie wyboru między licznymi opcjami menu jest wiedza o metodzie elementów skończonych, bowiem w każdej z tych grup zagadnień można popełnić błędy. Ważną informacja jest, że twórcy systemu ułatwiają ten wybór przez ustawienie opcji i parametrów domyślnych. Niestety, zwykle – nawet w prostych przypadkach – wyboru powinien dokonać użytkownik. Zaczyna się to już przy typie elementu skończonego, a kończy na odpowiednich równaniach konstytutywnych. Rozdział ten jest adresowany do potencjalnego lub faktycznego użytkownika takiego oprogramowania, który zna podstawowe zasady metod obliczeniowych, i jest w stanie zrozumieć jej zasadnicze przesłanie. Poniższy rozdział zawiera zadania wprowadzające studentów w bardzo istotne kwestie związane z analizą układów złożonych (złożenia). Analiza obiektów składających się z kilku części jest istotnym problem w procesie modelowania i symulacji mechanizmów i konstrukcji. W przypadku analizy złożeń konieczne jest zdefiniowanie charakteru współpracy elementów przez odpowiednie wprowadzenie zestawów kontaktowych. Rozdział został podzielony na trzy części: wprowadzenie (dla prowadzącego), przykład testowy (dla studentów), zadanie dodatkowe do samodzielnego wykonania (dla studentów). Tematyka ta przeznaczona jest na kilka zajęć laboratoryjnych.

1 Cel zadania

Student po wykonaniu zadań będzie umiał rozwiązać problemy związane z analizą statyczną złożeń części, utworzonych w systemie CAD. Pierwszym celem będzie zdobycie umiejętności poprawnego importu geometrii z zewnętrznego systemu CAD do oprogramowania SimCenter Femap. Geometria będzie podstawą do utworzenia modelu dyskretnego przez tworzenie siatki elementów skończonych. Student będzie potrafił zdefiniować kontakty na poszczególne części w złożeniu. Po przeprowadzonej analizie będzie potrafił zinterpretować uzyskane wyniki.

1.1 Wymagane umiejętności

- Student przed przystąpieniem do zajęć musi posiadać umiejętność modelowania przestrzennego w jednym z programów CAD (preferowany jest **SolidWorks lub Inventor**)

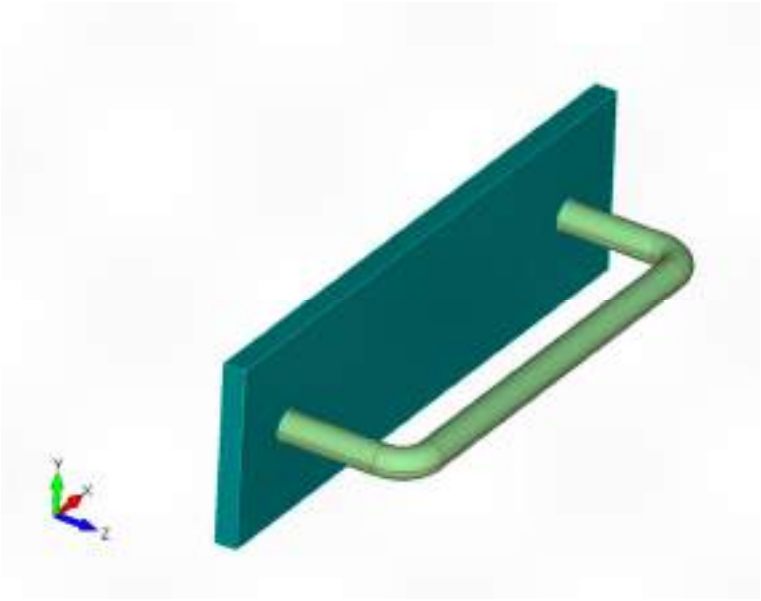
1.2 Zdobyte umiejętności

- Współpraca pomiędzy oprogramowaniem CAD a MES.
- Definiowanie warunków brzegowych na geometrii
- Definiowanie zestawów kontaktowych pomiędzy częściami w złożeniu - uwzględnienie układu z tarciem oraz części nieruchomych względem siebie
- Modyfikacja zewnętrznej bryły CAD

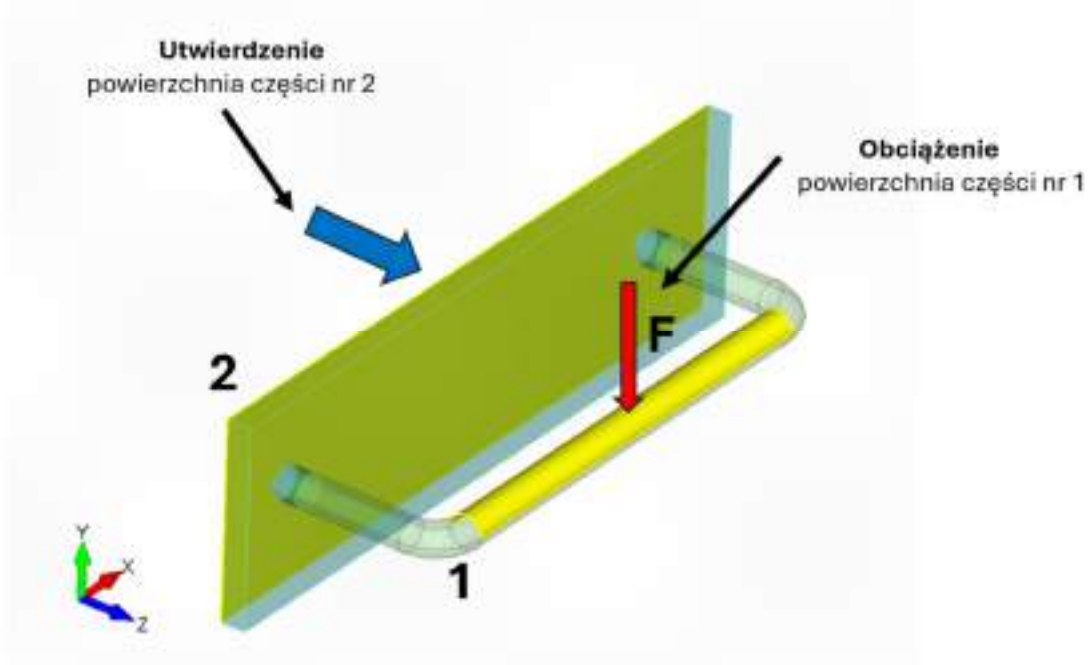
1.3 Zadanie do rozwiązania

Celem zadania jest analiza statyczna ugięcia statycznego złożenia przedstawionego na rysunkach 1 i2. Zostaną przeanalizowane trzy przypadki:

- Konstrukcja bez kontaktów - dwie części potraktowane jako jedna
- Części konstrukcji są połączone na ze sobą (brak wzajemnego ruchu)
- Części konstrukcji są ze sobą połączone umożliwiając ruch względem siebie



Rys. 1. Geometria (zadanie) do analizy

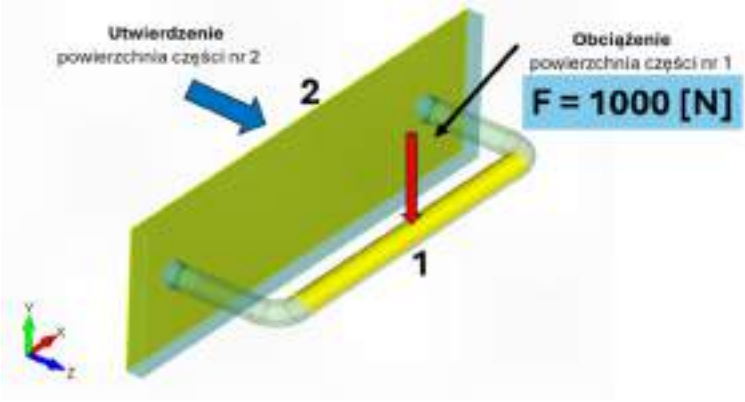


Rys. 2. Warunki brzegowe i obciążenie do przeprowadzenia analizy

2 Zadanie - Analiza złożenia CAD - kontakty

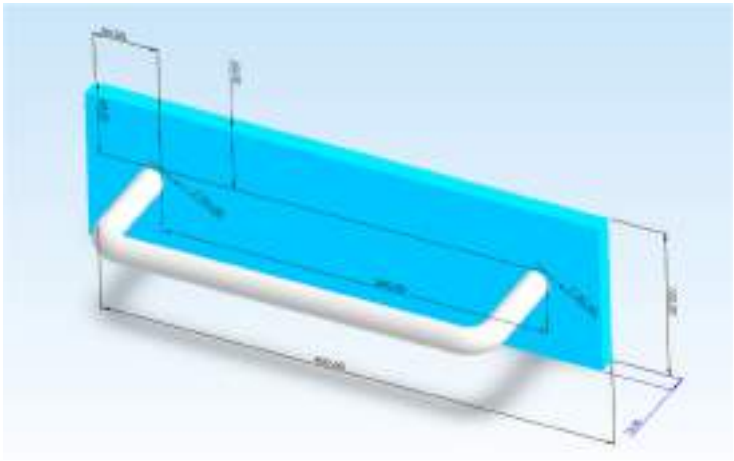
Celem tego ćwiczenia jest wyznaczenie ugięcia statycznego złożenia przedstawionego na poniższym rysunku. Do pręta przyłożona jest siła na powierzchni czołowej dłuższego końca, o wartości: $F = 1000[N]$. Przeanalizować trzy przypadki:

- konstrukcja bez kontaktów - dwie części zasymulowane jako jedna,
- części konstrukcji są połączona na stałe ze sobą - nieruchome połączenie,
- części konstrukcji są ze sobą połączone umożliwiając wzajemny ruch.



Rys. 3. Warunki symulacji uchwytu

Wymiary zadania są przedstawione na rysunku poniżej (rys. 4).



Rys. 4. Ogólne wymiary geometryczne modelu do analizy

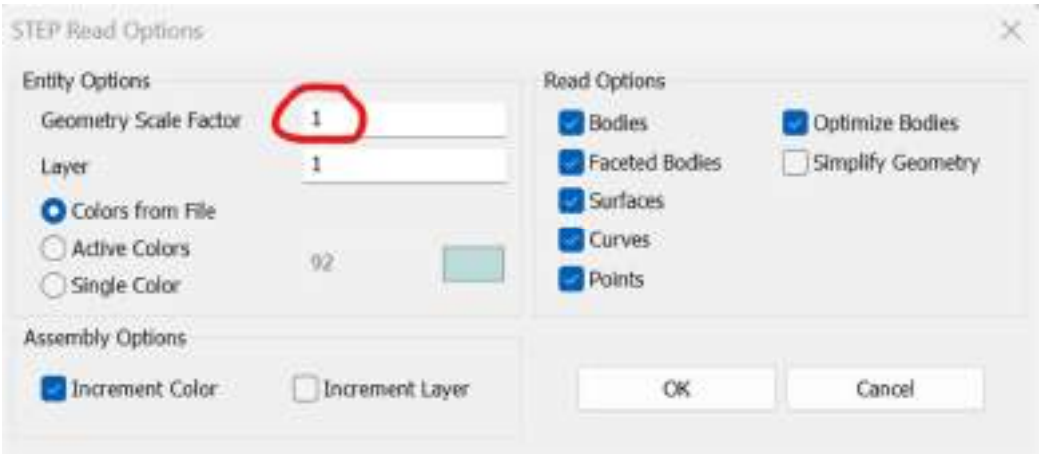
Materiał – stal - o parametrach:

- moduł Young'a - $E = 2,1 \cdot 10^{11}[Pa]$,
- liczba Poissona - $\nu = 0,3$,
- gęstość - $\rho = 7800[\frac{kg}{m^3}]$

2.1 Import geometrii

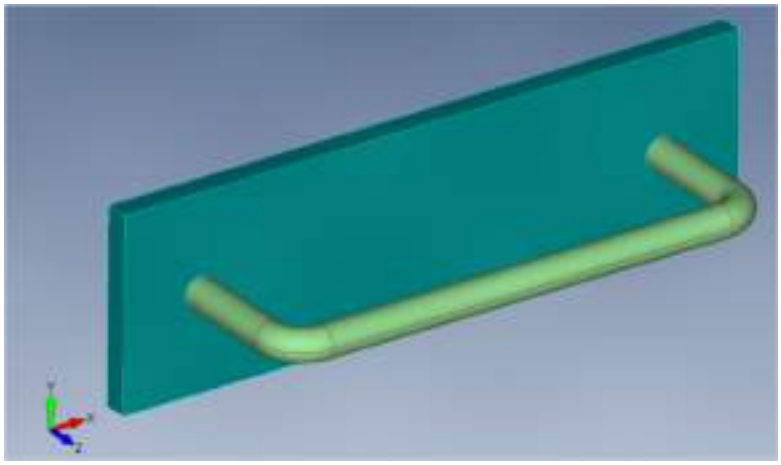
UWAGA: Geometria do importu powinna być wyeksportowana z programu SolidWorks za pośrednictwem formatu STEP (.STEP)*

W celu poprawnego wczytania geometrii do programu Femap należy zmienić do- myślne ustawienia jednostek w jakich model CAD zostanie zaimportowany. W tym celu w programie **Simcenter Femap** wybierz **File**, **Import -> Gemetry** (wskaż zapisaną geometrię w formacie *STEP (*.STEP))* a następnie w oknie **STEP Read Options** (rys. 5) przy **Solid Geometry Scale Factor** wpisz wartość *1.* Kliknij **OK**.



Rys. 5. Ustawienie poprawnej skali podczas importu geometrii do oprogramowania Simcenter Femap

Zaimportowana geometria powinna wyglądać następująco (rys 6.)



Rys. 6. Poprawnie zaimportowana geometria w Simcenter Femap

Zapisz model

poprzez **File**, **Save**.

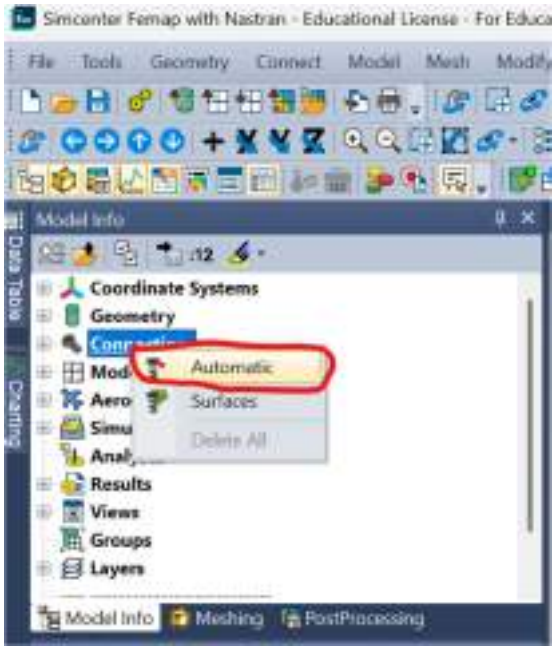
2.2 Definicja zestawów kontaktowych dla części współpracujących - połączenie nieruchome

Kontakty są definiowane w celu ograniczenia stopni swobody dla części, które nie będą utwierdzone. Tworzenie kontaktów polega na wprowadzeniu trzech parametrów:

- Określenie typu połączenia (ang. *Connection Property*) - określenie warunków kontaktu potrzebnych do zdefiniowania żądanego zachowania się pary współpracujących regionów;
- Określenie współpracujących regionów (ang. *Connection Regions*) - określenie na modelu elementów, pomiędzy którymi dochodzi do interakcji. Mogą to być elementy geometrii tj. powierzchnie, krzywe lub punkty;
- Określenie właściwości połączenia (ang. *Connectors*) na podstawie powyższych parametrów. Definiowana jest zależność *Master* - *Slave* pomiędzy elementami współpracujących.

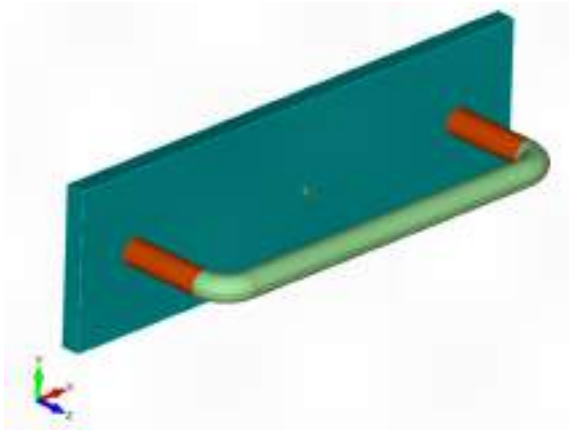
Kontakty można również definiować ręcznie lub w sposób automatyczny na pod- stawie maksymalnej odległości pomiędzy częściami w złożeniu. W tym zadaniu zostanie przedstawiony automatyczny sposób tworzenia kontaktów.

Z drzewa modelu **Model Info** (rys. 7.), kliknij prawym przyciskiem myszy (PPM) na opcję **Connections** i **Automatic**. W oknie **Entity Selection - Select Solid(s) to Detect Connections** kliknij **Select All** i **OK**. W oknie **Auto Detection Options for Connections** w polu **Connection Property** zaznacz **Glued**, kliknij **OK** i **Cancel**.



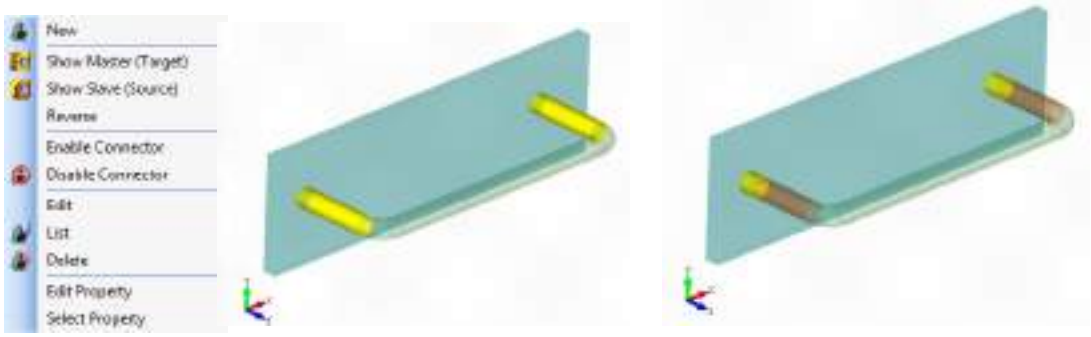
Rys. 7. Panel model info ->>>

Utworzony zostanie nowy region (rys. 8.), który zostaje dodatkowo wyróżniony na ekranie podglądu modelu kolorem pomarańczowym.



Rys. 8. Znaleziony region dla zestawów kontaktowych

Zależność pomiędzy regionami dla każdego kontaktu można dodatkowo oddzielnie przedstawić rozwijając drzewo ustawień przy **Connectors** i klikając PPM na *1..Region 1-2* i z menu kontekstowego wybierając **Master** lub **Slave**. W celu powrotu do poprzedniego stanu wyświetlania należy wybrać **Ctrl+G**.



(a) (b) Master (c) Slave

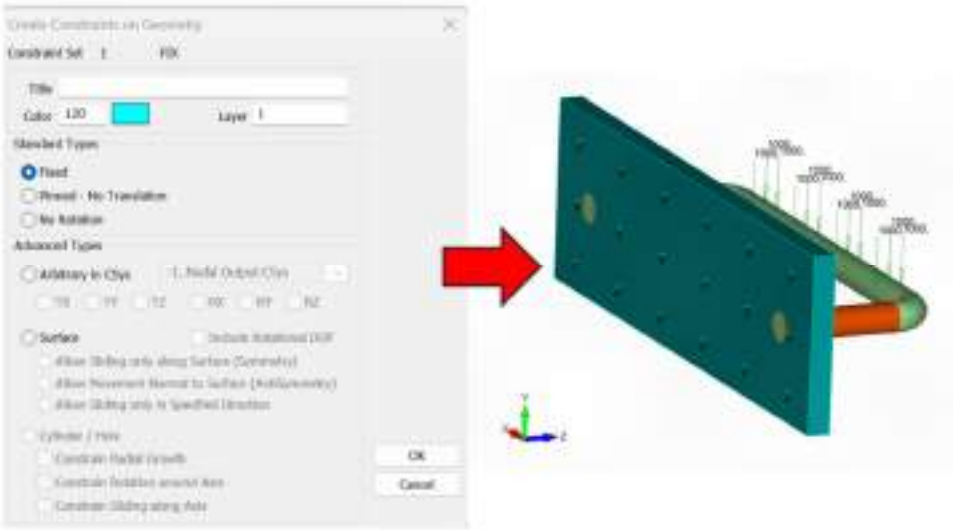
Rys. 9. Zestawy kontaktowe w badanej geometrii

2.3 Przypisanie warunków brzegowych

W celu przyłożenia obciążenia na dłuższy koniec pręta wybierz z menu głównego **Model**, **Load**, **On Surface...**. W oknie **New Load Set** wprowadź nazwę, np. *Sila*, kliknij **OK**. W oknie **Entity Selection - Select Surface(s) to Select** wybierz na ekranie właściwą powierzchnię, kliknij **OK**. W oknie **Create Load on Surface** wybierz typ obciążenia **Force**. W polu **Load** wprowadź wartość siły Fy = -1000[N] (rys. 10), Kliknij **OK** i **Cancel**.

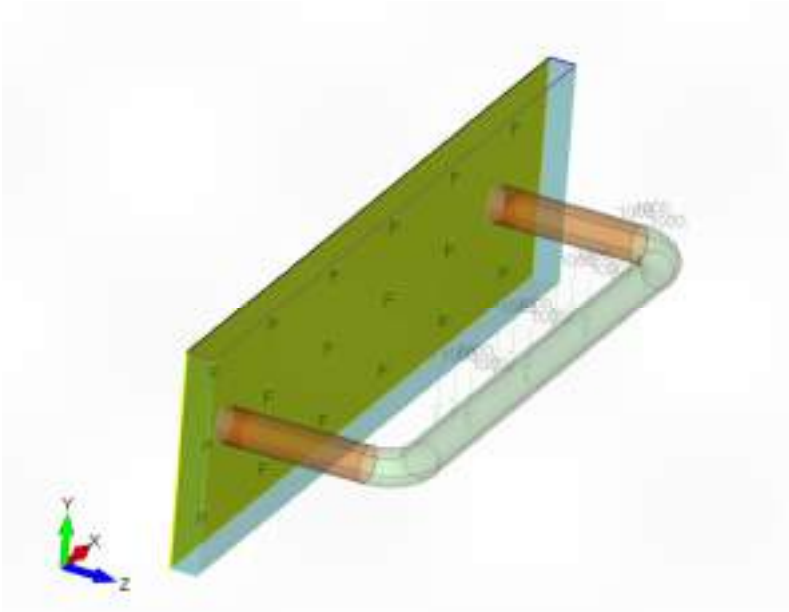
Rys. 10. Definicja obciążenia dla części nr 1 – widok zadanej poprawnie siły

Następnie ponownie wybierz z menu głównego **Model**, **Constraint**, **On Surface**. W **On Constraint Set** w polu **Title** wprowadź nazwę, np. *Utwierdzenie* - **OK**. W oknie **Entity Selection - Select Surface(s) to Select** wskaż na ekranie dwie powierzchnie na kwadratowej części - prostopadłe do przyłożonej siły, kliknij **OK**. W oknie **Create Constraints on Geometry** w polu **Standard Types** wybierz opcję *Fixed* (rys. 11), kliknij **OK** i **Cancel**.



Rys. 11. Definicja warunków brzegowych – widok poprawnie zadanego utwierdzenia

Model powinien wyglądać następująco (rys. 12):

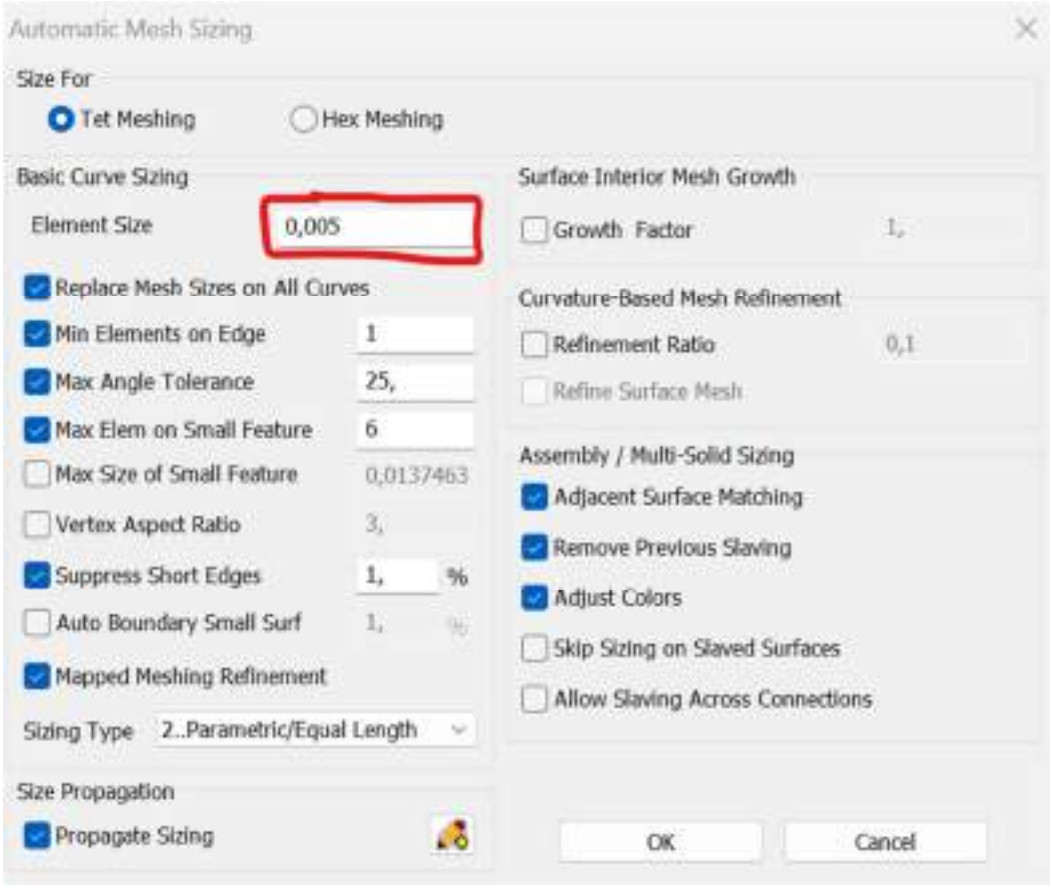


Rys. 12. Geometria z zadanymi warunkami brzegowymi i obciążeniem

Zapisz model poprzez **File**, **Save**.

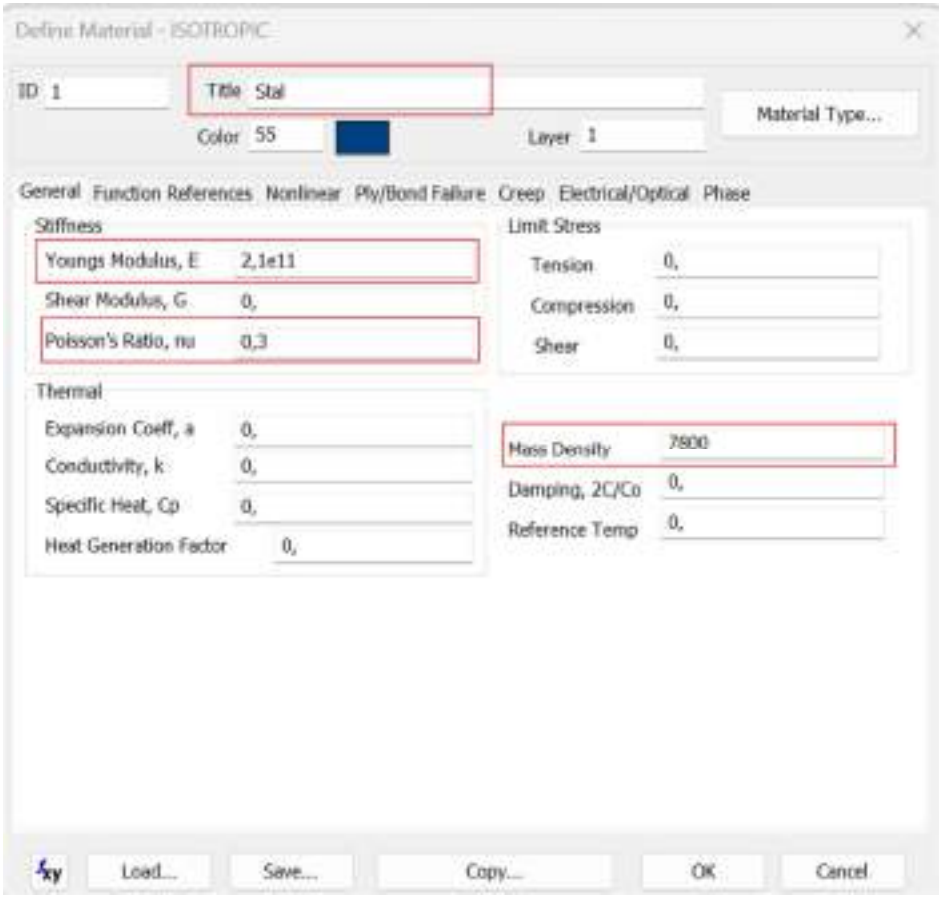
2.4 Budowa modelu numerycznego – dyskretyzacja MES

W celu określenia wielkości elementu objętościowego wybierz **Mesh**, **Mesh Control**, **Size On Solid....** W oknie **Entity Selection - Select Solid(s) to Set Mesh Size** kliknij **Select All** i **OK**. W oknie **Automatic Mesh Sizing** w polu **Size For** zaznacz opcję **Tet Meshing** w celu wprowadzenia ustawień dla elementów trójwymiarowych typu TETRA; następnie w polu **Basic Curve Sizing** przy **Element Size** wprowadź wartość 0, 005[*m*] (rys. 13). Kliknij **OK**.



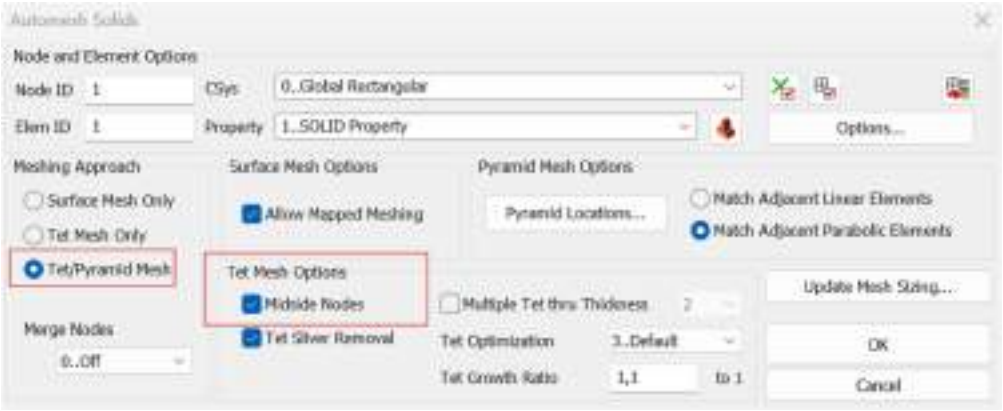
Rys. 13. Definicja rozmiaru globalnego elementu dla części 1 i 2

Ponownie wybierz **Mesh**, **Geometry**, **Solids** - kliknij **Select All** i **OK**. W oknie **Define Material - ISOTROPIC** w polu **Title** wprowadź nazwę, np. *Stal* oraz zadane właściwości materiałowe (rys. 14). Kliknij **OK**.



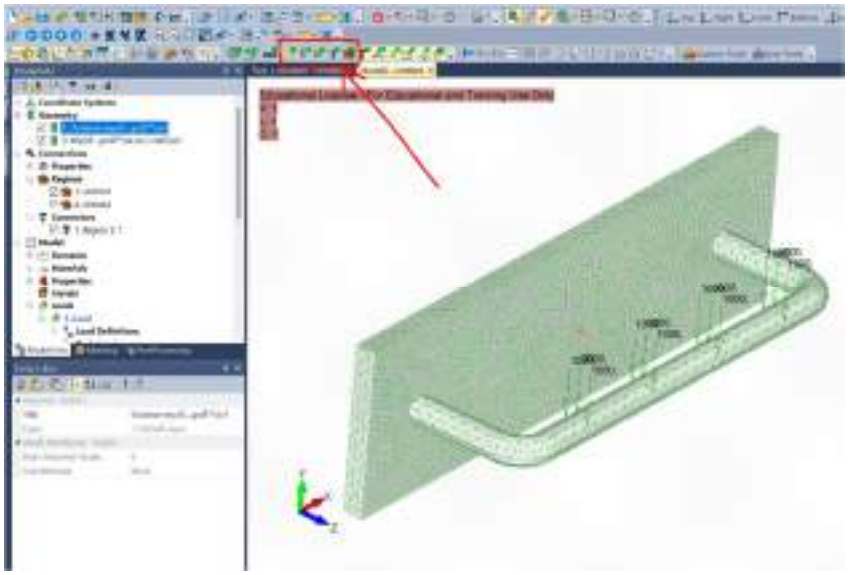
14. Parametry materiału

Następnie w oknie **Automesh Solids** kliknij **OK** pozostawiając domyślne ustawienia. Zostawiając zaznaczoną opcje *Midside Nodes* w polu **Mesh Generation** zostanie utworzona siatka elementów skończonych drugiego rzędu (rys. 15)



Rys. 15. Opcje okna generacji siatki przestrzennej

Na pasku **Entity Display** odznacz wyświetlanie punktów, krzywych, powierzchni oraz regionów kontaktów. Model powinien wyglądać następująco:



Rys. 16. Wyświetlenie modelu obliczeniowego (wyłączenie geometrii)

Zapisz model poprzez **File**, **Save**.

2.5 Analiza numeryczna – proces obliczeniowy

Z menu głównego wybierz **Model**, **Analysis.**

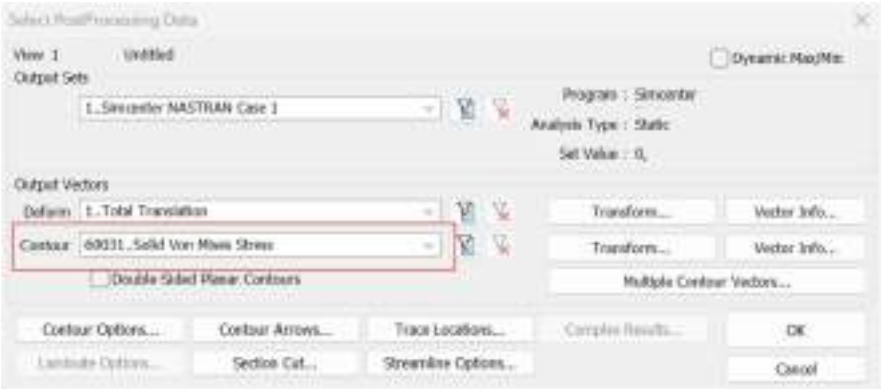
W oknie **Analysis Set Manager** kliknij **New...**, a następnie w oknie **Analysis Set** w polu **Title** wprowadź nazwę analizy, np. *Zlozenie* z menu **Analysis Program** wybierz *36 .Simcenter Nastran*, a z menu **Analsis Type** wybierz *1..Static*. Kliknij **OK** i **Analyze**.

Po poprawnie ukończonych obliczeniach w panelu **SimCenter Nastran Analysis Monitor** jest linia *Analysis complete* - czwarta od dołu okna z logiem (rys. 17).

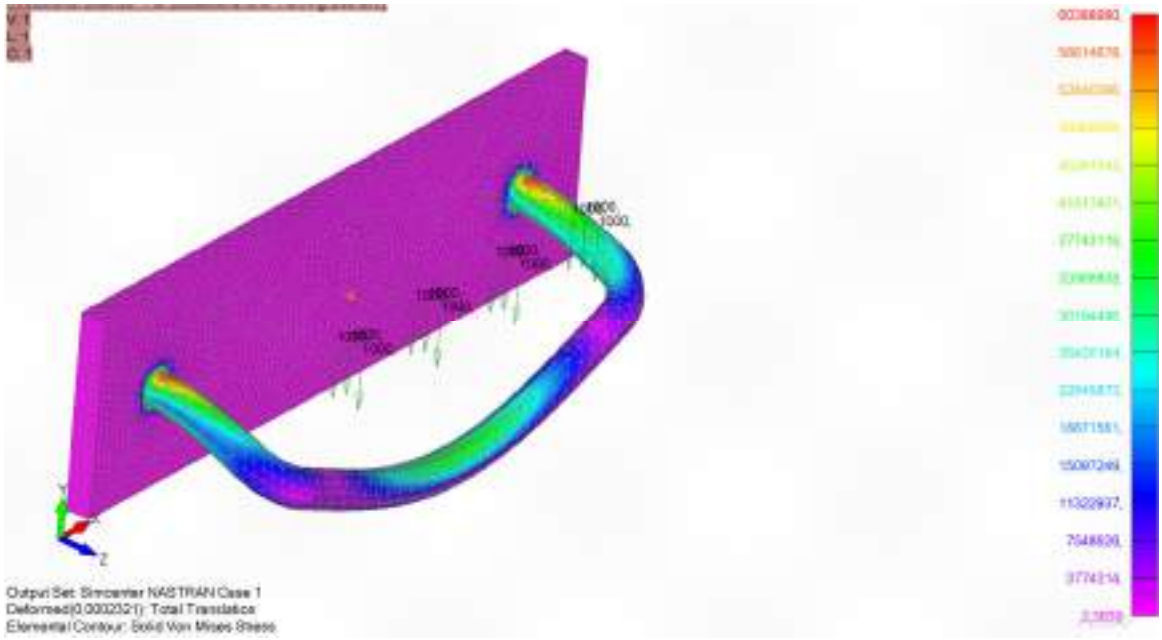
Rys. 17. Widok loga solwera – proces u obliczeniowego →

2.6 Obróbka wyników

Z menu głównego wybierz **View**, **Select**. W oknie **View Select** z pola **Deformed Style** zaznacz opcję *Deform*, w polu **Contour Style** zaznacz opcję *Contour*. Kliknij **Deformed and Contour Data...**. W oknie **Select PostProcessing Data** w polu **Output Vectors** z menu **Deformation** wybierz *1..Total Translation*, z pola **Contour** wybierz *60031..Solid Von Mises Stress* - kliknij dwa raz **OK** (rys. 18, 19).

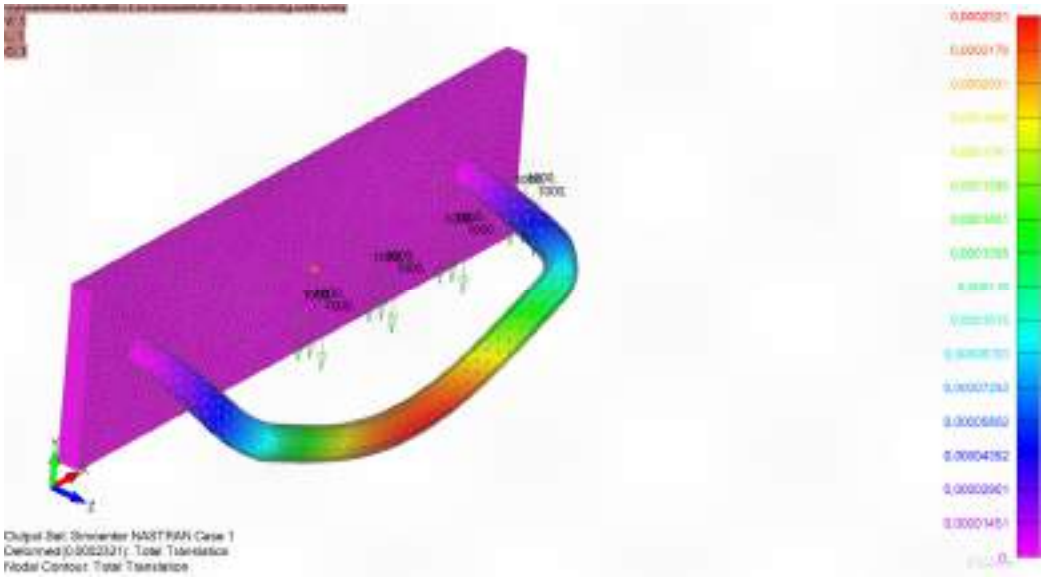


Rys. 18. Ustawienia wizualizacji wyników – naprężenia zredukowane



Rys. 19. Rozkład naprężeń zredukowanych według Hubera-Misesa

Kliknij na przycisk **F5** z klawiszy funkcyjnych. Kliknij **Deformed and Contour Data...**, z pola **Contour** wybierz *3..T2 Translation*.



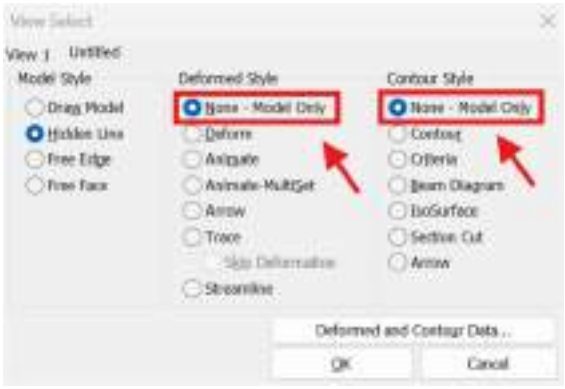
Rys. 20. Rozkład przemieszczeń wypadkowych

Zanotuj maksymalne naprężenia zredukowane Hubera-Misesa w[*MPa*], i maksymalne przemieszczenie w kierunku *Y* w [*mm*]. Wartości te będą konieczne dla porównania z kolejnym przypadkiem.

Zapisz model poprzez **File**, **Save**.

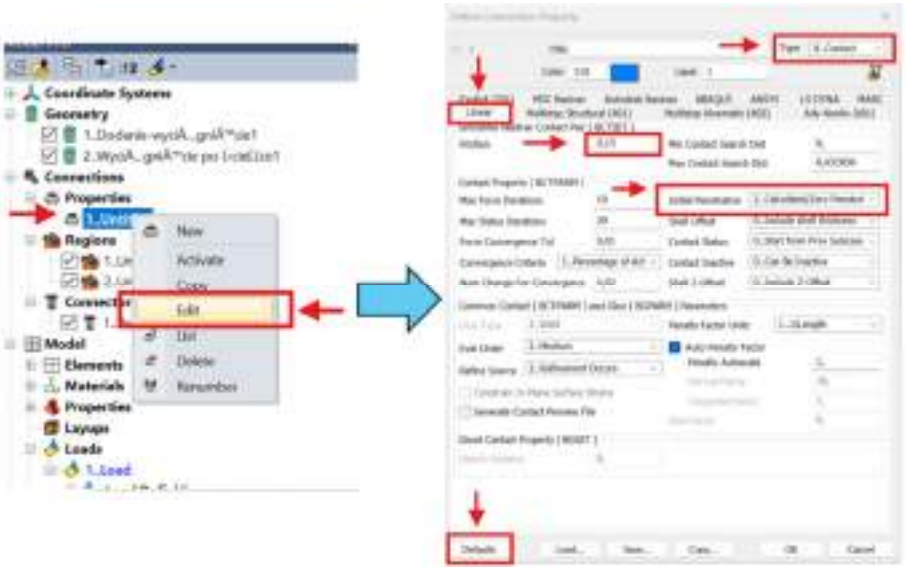
2.7 Modyfikacja zadania – zmiana zestawu kontaktowego - przypadek połączenia umożliwiający wzajemny ruch

Z menu głównego wybierz **View**, **Select**. W oknie **View Select** z pola **Deformed Style** zaznacz opcję *None - Model Only*, w polu **Contour Style** zaznacz opcję *None - Model Only* (rys. 21). Kliknij **OK**.



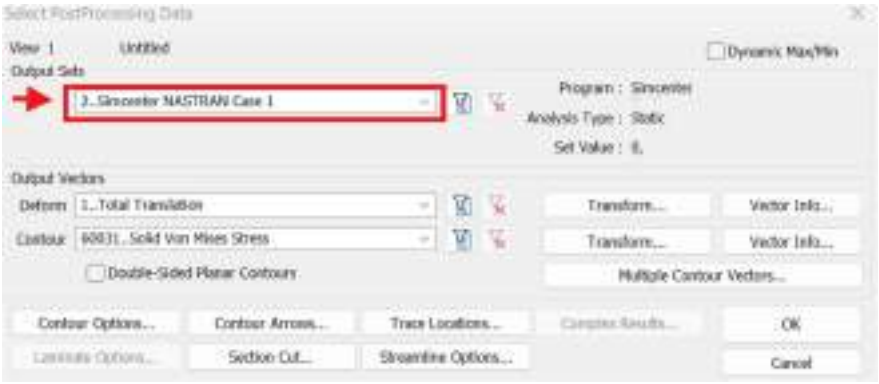
Rys. 21. Widok okna „View Select"

W panelu/oknie **Model Info** rozwiń drzewo opcji przy **Connections**, **Properties**. Kliknij PPM na *1..Untitled*. W oknie **Define Connection Property** z menu **Connect Type** wybierz *0..Contact*. Kliknij zakładkę **Linear** w celu wprowadzenia ustawień dla liniowego typu kontaktu. W polu **Contact Pair (BCTSET)** w polu **Friction** wprowadź współczynnik tarcia. W tym przypadku (stal–stal) wprowadź 0.15. Z pola **Contact Property (BCTPARM)** z menu **Initial Penetration** wybierz *2..Calculated/Zero Penetrations* (rys. 22).

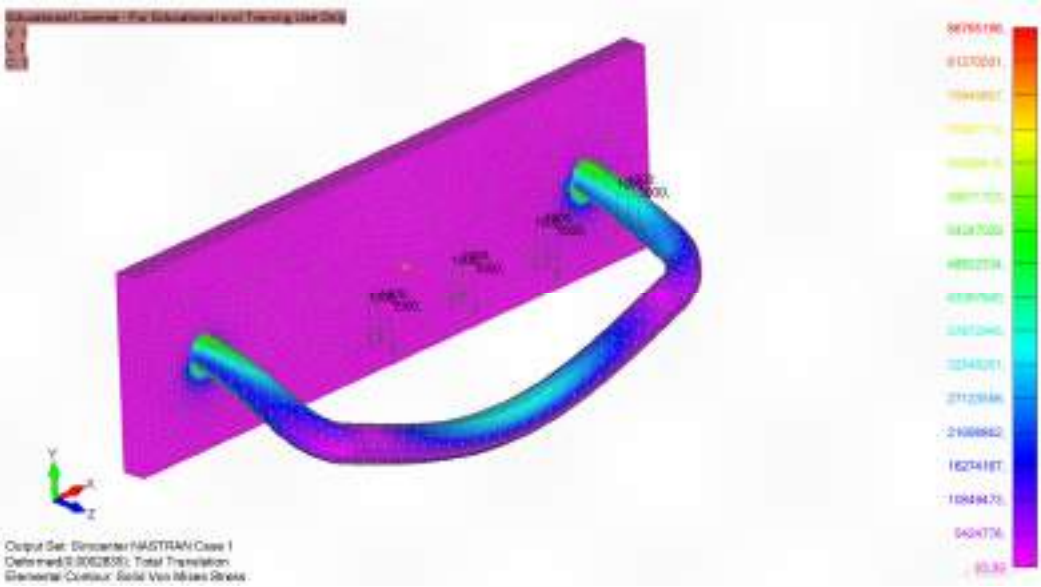


Rys. 22. Widok okna modyfikacji kontaktu pomiędzy współpracującymi elementami

Zapisz model i uruchom analizę. Następnie wyświetl wyniki tak jak opisano już wyżej. W celu wczytania wyników kolejnej analizy, w oknie **Select PostProcessing Data** w polu **Output Set** wybierz z menu *2..Simcenter NASTRAN Case 1* (rys. 23).

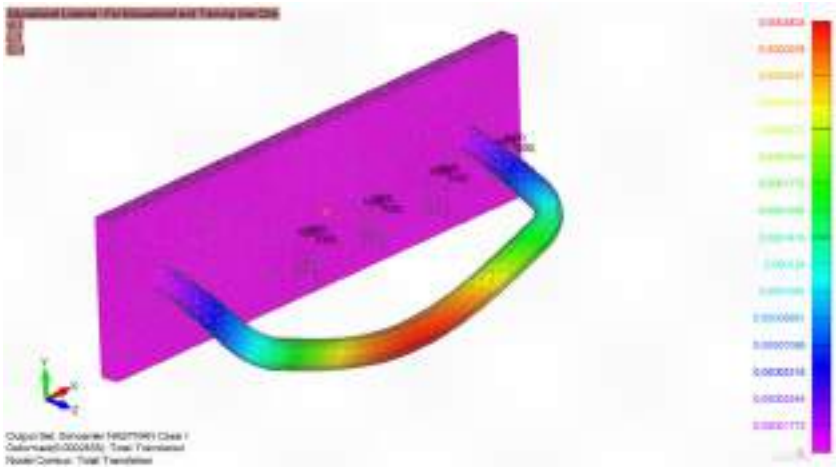


Rys. 23. Wczytanie wyników w drugiej analizy (z uwzględnieniem kontaktu).



Rys. 24. Rozkład naprężeń zredukowanych według Hubera-Misesa

Zanotuj, ile wynosi maksymalne naprężenie zredukowane według hipotezy Hubera-Misesa w [*MPa*], oraz maksymalne przemieszczenie w kierunku *Y* w [*mm*].



Rys. 25. Rozkład przemieszczeń wypadkowych

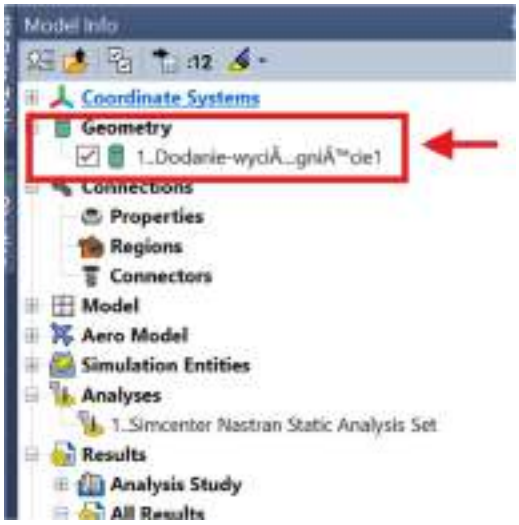
Dodatkowo można zaobserwować, że uchwyt wysunął się ze ścianki.

Zapisz model poprzez **File**, **Save**.

2.8 Nowe zadanie - jednolita część

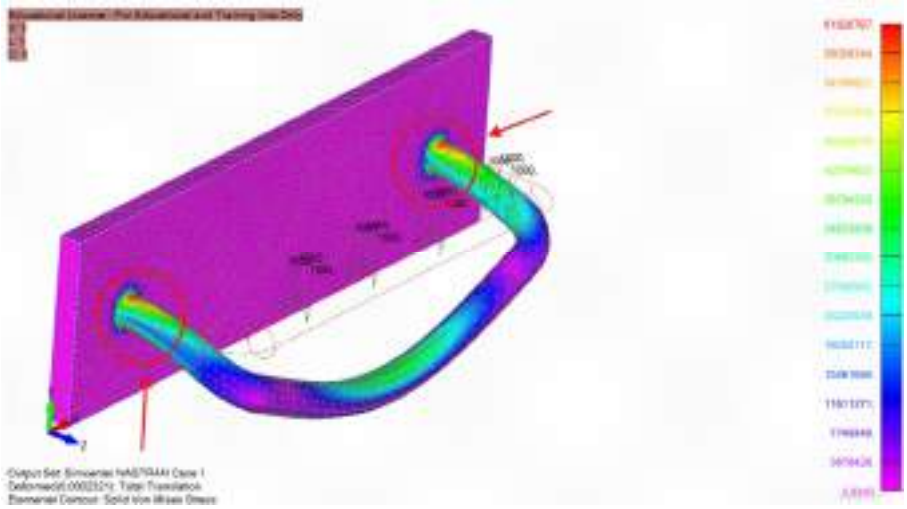
Utwórz nowy plik poprzez **File**, **New**.

Zaimportuj geometrię, podobnie jak wyżej. W celu połączenia dwóch części w jedną objętość wybierz **Geometry**, **Solid**, **Add....** W oknie **Entity Selection - Select Solid(s) to Add** kliknij **Select All** i **OK**. W panelu **Model Info** rozwiń drzewo ustawień przy **Geometry** - powinna być dostępna jedna (rys. 26).

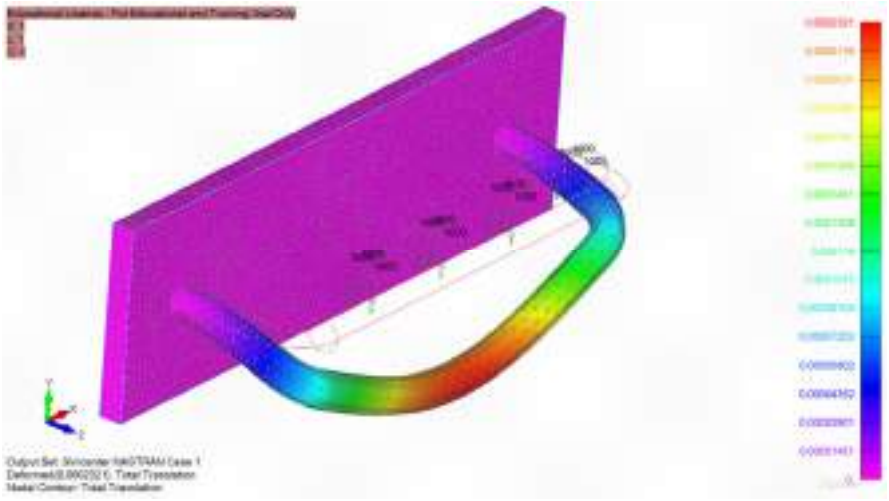


Rys. 26. Efekt po połączeniu geometrii (widoczna jedna bryła).

Dalej postępuj zgodnie z instrukcjami powyżej, z pominięciem etapu definicji zestawów kontaktowych - zdefiniuj warunki brzegowe oraz utwórz model dyskretny MES z takimi samymi ustawieniami jak poprzednio. Zapisz model pod inną nazwą i uruchom analizę. Następnie wyświetl wyniki – widok jak poniżej (rys. 27, 28).



Rys. 27. Rozkład naprężeń zredukowanych według Hubera-Misesa



Rys. 28. Rozkład przemieszczeń wypadkowych

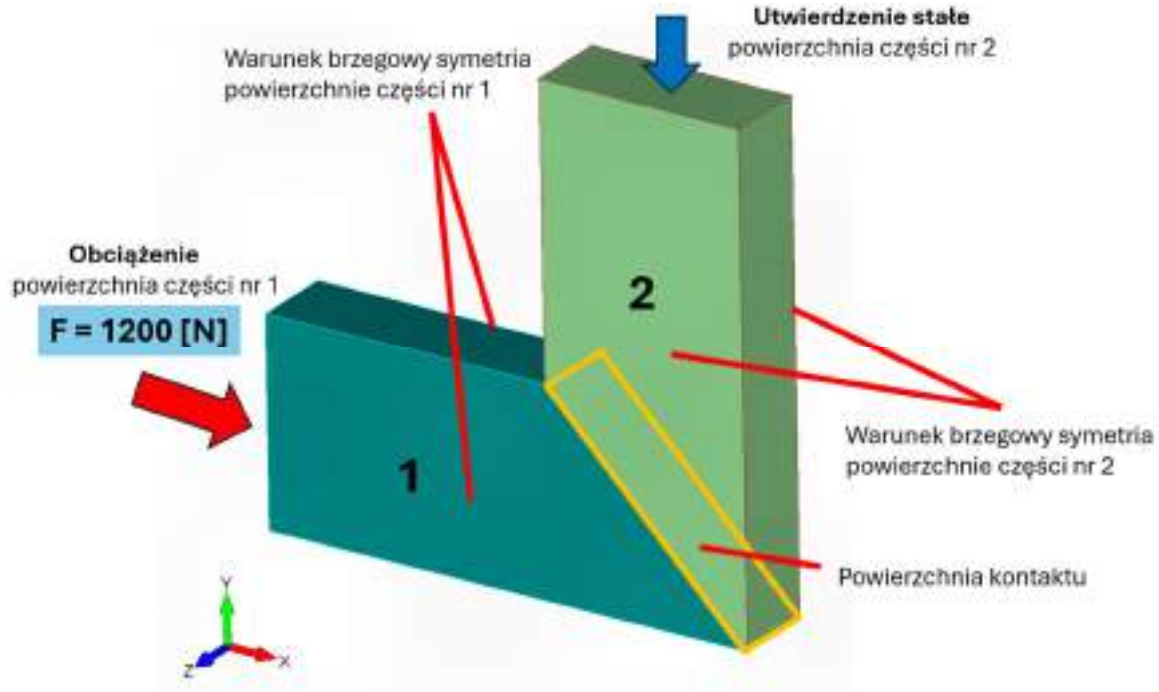
Zanotuj, ile wynosi maksymalne naprężenie zredukowane Hubera-Misesa w [*MPa*] oraz maksymalne przemieszczenie po kierunku *Y* w [*mm*].
Różnicę między rozwiązaniem numerycznym z kontaktami i bez nich należy wyznaczyć z zależności:

$$\Delta_{wz} = \frac{\sigma_{BezKontaktów} - \sigma_{ZKontaktami}}{\sigma_{BezKontaktów}} \cdot 100\% = \ldots\ldots\ldots\ldots\ldots \quad (1)$$

Wyznacz wartość:

3 Zadanie dodatkowe

Wyznaczyć ugięcie statyczne poniższej konstrukcji. Uwzględnić tarcie pomiędzy częściami. Przyłożona siła do podstawy wynosi $F - 1200[N]$



Rys. 29. Dodatkowe zadanie do samodzielnego rozwiązania

Geometria dostępna na e-kursie. Warunki brzegowe zgodnie z rysunkiem 29.

Dane materiałowe o parametrach:

- moduł Young'a - $E = 2,1 \cdot 10^{11}[Pa]$,
- liczba Poissona - $\nu = 0,3$,
- gęstość - $\rho = 7800[\frac{kg}{m^3}]$
- Współczynnik tarcia 0.15

Podsumowanie

Przedstawiony rozdział poświęcony został istotnemu zagadnieniu związanemu z importem i analizą geometrii złożenia z systemu CAD do oprogramowania MES z uwzględnieniem zestawów kontaktowych. Proces ten bardzo często jest niezbędny w przypadku analizy złożeń i mechanizmów, gdzie należy zdefiniować charakter współpracy poszczególnych części. Poprawna definicja kontaktów pozwala uzyskać właściwe rozwiązanie.

Rozdział pokazuje różne podejścia do analizy złożeń w systemie MES i związane z tym różne rozwiązania pokazujące wpływ definicji zestawu kontaktu na wynik. Proces definicji współpracujących ze sobą elementów pozwala w lepszy sposób zrozumieć istotę modelowania problemów rzeczywistych z wykorzystaniem modeli dyskretnych, który spełnią wymagania postawione przez narzędzia obliczeniowe. Informację zawarte w rozdziale zawierają praktyczną wiedzę, która pozawala w sposób efektywny zrozumieć rolę kontaktów w modelowaniu obiektów złożonych.

Bibliografia

Domański, J., SolidWorks. Projektowanie maszyn i konstrukcji, Wydawnictwo Helion, Gliwice 2017.

Fortuna Z., Macukow B., Wąsowski J., Metody numeryczne, WNT, Warszawa 1993.

Rakowski G., Kacprzyk Z., Metoda elementów skończonych w mechanice konstrukcji, OWPW, Warszawa 2016.

Zienkiewicz O. C., Metoda elementów skończonych, Arkady, Warszawa 1972.

Sztuczna inteligencja w automatyce i robotyce

dr Paweł Sobczak

1. Algorytmy genetyczne i ewolucyjne

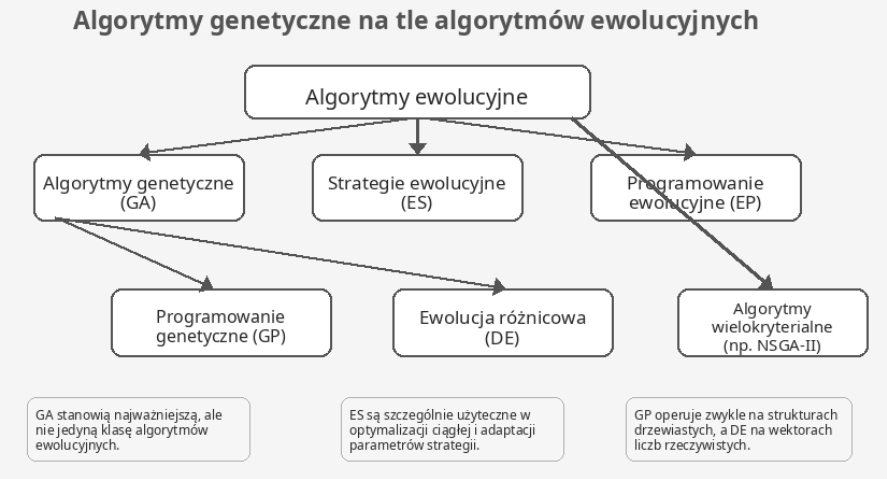
1.1. Wprowadzenie

Algorytmy ewolucyjne stanowią rodzinę metod obliczeniowych inspirowanych mechanizmami doboru naturalnego, dziedziczenia i zmienności biologicznej. Ich wspólną ideą jest przeszukiwanie przestrzeni rozwiązań przez operowanie na populacji kandydatów, a nie na pojedynczym punkcie, jak ma to miejsce w wielu klasycznych metodach optymalizacji. W każdej iteracji, nazywanej zwykle pokoleniem, rozwiązania są oceniane, lepsze osobniki uzyskują większą szansę przekazania informacji do następnej generacji, a operatory rekombinacji i mutacji wytwarzają nowe warianty. Z perspektywy praktyki inżynierskiej nie jest to wyłącznie metafora biologiczna. Jest to dobrze ugruntowany sposób projektowania procedur optymalizacyjnych, który pozwala skutecznie radzić sobie z funkcjami celu trudnymi, kosztownymi obliczeniowo, nieliniowymi, nieciągłymi albo wielokryterialnymi.

Za klasyczne źródła tej dziedziny uznaje się prace Johna Hollanda[1], który sformalizował algorytm genetyczny, oraz rozwój pokrewnych nurtów badawczych powstających równolegle w Stanach Zjednoczonych i w Europie. W literaturze podkreśla się jednak, że algorytm genetyczny nie wyczerpuje całego pola metod ewolucyjnych. Jest on jedną z najważniejszych, historycznie najbardziej rozpoznawalnych klas, lecz obok niego rozwinęły się także strategie ewolucyjne, programowanie ewolucyjne, programowanie genetyczne, a później również ewolucja różnicowa i liczne algorytmy wielokryterialne. Dla studenta kierunku automatyki i robotyki rozróżnienie to ma znaczenie praktyczne, ponieważ dobór właściwej odmiany algorytmu zależy od typu zadania: inaczej stroi się parametry regulatora, inaczej optymalizuje strukturę programu sterowania, a inaczej wyznacza kompromis między energią, dokładnością i czasem ruchu robota.

[1] J. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.

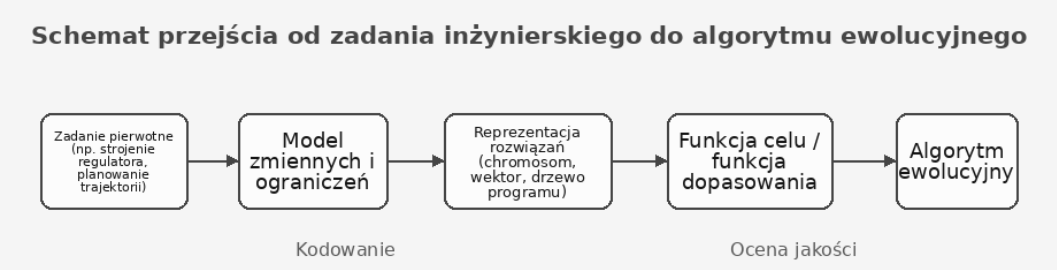
W opisie tego rozdziału rozróżnia się dwa poziomy (podejścia) do algorytmów genetycznych i ewolucyjnych (rys. 1): poziom pierwszy dotyczy ogólnej rodziny algorytmów ewolucyjnych, a więc metod opartych na populacji, selekcji i zmienności, poziom drugi dotyczy szczególnego przypadku, jakim jest algorytm genetyczny. Taki układ jest bardziej zgodny z literaturą akademicką i pomaga uniknąć częstego nieporozumienia polegającego na traktowaniu wszystkich algorytmów ewolucyjnych jakby były po prostu odmianami jednego klasycznego AG.



Rys. 1. Algorytmy genetyczne jako jedna z klas algorytmów ewolucyjnych

Źródło: Opracowanie własne.

Rys. 2. przedstawia kolejne etapy przejścia od praktycznego zadania (problemu) inżynierskiego do zastosowania algorytmu ewolucyjnego. Proces zaczyna się od zdefiniowania problemu pierwotnego, następnie przechodzi przez model zmiennych i ograniczeń oraz reprezentację rozwiązań, czyli sposób zapisania kandydatów na rozwiązania. Dalej określana jest funkcja celu lub funkcja dopasowania, która pozwala ocenić jakość rozwiązań, a na końcu uruchamiany jest właściwy algorytm ewolucyjny.



Rys. 2. Schemat przejścia od zadania inżynierskiego do algorytmu ewolucyjnego

Źródło: Opracowanie własne.

1.2. Algorytmy genetyczne a algorytmy ewolucyjne - relacja pojęć

Algorytmy ewolucyjne to nadrzędna kategoria metod optymalizacyjnych i przeszukujących, które wykorzystują populację rozwiązań, funkcję oceny oraz mechanizmy selekcji i zmienności. W najprostszym ujęciu można powiedzieć, że każdy algorytm ewolucyjny realizuje ten sam schemat myślenia: generuje zbiór kandydatów, ocenia ich jakość, preferuje lepszych i wytwarza nowe rozwiązania na podstawie już znalezionych. Różnice między poszczególnymi odmianami dotyczą przede wszystkim reprezentacji rozwiązania, sposobu tworzenia potomstwa, znaczenia mutacji i rekombinacji oraz reguł zastępowania populacji.

Algorytm genetyczny jest zatem szczególnym przypadkiem algorytmu ewolucyjnego. W tradycyjnym ujęciu operuje on na chromosomach zapisanych jako ciągi symboli, najczęściej binarnych, stosuje selekcję preferującą lepiej przystosowanych osobników, a nową populację tworzy głównie przez krzyżowanie i mutację. Historycznie właśnie ten model stał się najbardziej znanym wzorcem dydaktycznym, ponieważ bardzo czytelnie pokazuje, jak zakodować zmienne, jak zdefiniować funkcję dopasowania i jak krok po kroku przebiega sztuczna ewolucja.

Strategie ewolucyjne wyrosły z potrzeb optymalizacji parametrów ciągłych i od początku większy nacisk kładły na mutację niż na krzyżowanie. Ważną ich cechą jest możliwość adaptowania nie tylko samych rozwiązań, lecz także parametrów sterujących procesem mutacji. Programowanie ewolucyjne historycznie koncentrowało się na ewolucji zachowania i automatów stanów, a programowanie genetyczne przeniosło ideę ewolucji na struktury programów, najczęściej reprezentowane jako drzewa składniowe. Ewolucja różnicowa z kolei posługuje się prostymi, lecz bardzo skutecznymi operatorami na wektorach liczb rzeczywistych i dzięki temu bywa wyjątkowo konkurencyjna w ciągłej optymalizacji inżynierskiej.

To rozróżnienie nie jest jedynie porządkiem terminologicznym. W praktyce oznacza ono, że nie istnieje jeden uniwersalny przepis na algorytm ewolucyjny. Ta sama idea populacyjnego przeszukiwania może zostać zaimplementowana na wiele sposobów, a skuteczność zależy od zgodności reprezentacji i operatorów z naturą problemu. Właśnie dlatego w dalszej części rozdziału klasyczny AG będzie traktowany jako punkt wyjścia do zrozumienia całej rodziny metod, a nie jako jedyny model ewolucyjnego obliczania.

1.3. Dlaczego metody ewolucyjne są potrzebne w automatyce i robotyce

W automatyce i robotyce problemy optymalizacyjne bardzo rzadko mają postać wygodnych zadań podręcznikowych. Często są obciążone ograniczeniami, obejmują wiele zmiennych, wykorzystują kosztowne symulacje i prowadzą do funkcji celu, które są wielomodalne, nieregularne, a nawet częściowo nieciągłe. Przy strojeniu regulatora, projektowaniu trajektorii, doborze parametrów filtracji, rozmieszczeniu czujników albo syntezie reguł sterowania nie zawsze dysponuje się pochodnymi, a jeśli nawet są dostępne, to mogą prowadzić do lokalnych ekstremów odległych od rozwiązania rzeczywiście najbardziej użytecznego.

Właśnie w takich sytuacjach metody ewolucyjne okazują się szczególnie atrakcyjne. Nie wymagają ścisłej wiedzy analitycznej o funkcji celu, dobrze współpracują z modelami symulacyjnymi i pozwalają oceniać rozwiązania na podstawie wyniku eksperymentu numerycznego lub rzeczywistego testu stanowiskowego. Innymi słowy, algorytmowi nie trzeba tłumaczyć, jak policzyć gradient, jeżeli wystarczy wiarygodnie ocenić jakość kandydata. Dla inżyniera jest to bardzo ważna własność, ponieważ wiele praktycznych układów sterowania i systemów robotycznych zachowuje się zbyt złożenie, aby wygodnie stosować klasyczne techniki oparte na silnych założeniach gładkości i wypukłości.

Drugim argumentem przemawiającym za metodami populacyjnymi jest ich naturalna zdolność do równoległego badania wielu rejonów przestrzeni rozwiązań. W zadaniach z licznymi ekstremami lokalnymi, typowych choćby dla strojenia złożonych układów nieliniowych, pojedynczy punkt startowy może bardzo łatwo doprowadzić procedurę do niepożądanego optimum lokalnego. Populacja rozwiązań redukuje to ryzyko, ponieważ jednocześnie utrzymuje kilka hipotez o tym, gdzie warto prowadzić dalsze poszukiwania. Nie jest to gwarancja sukcesu, ale jest to przewaga strukturalna nad metodami lokalnymi.

Trzecią cechą istotną z punktu widzenia robotyki jest łatwość formułowania kryteriów wielokryterialnych. W praktyce nie optymalizuje się zwykle jednego parametru. Chce się jednocześnie skrócić czas przejazdu, zmniejszyć zużycie energii, poprawić dokładność pozycjonowania, ograniczyć drgania i zachować zapas bezpieczeństwa. Algorytmy ewolucyjne idealnie pasują do tego podejścia, ponieważ potrafią zarządzać całą populacją różnorodnych rozwiązań jednocześnie. Zamiast zmuszać nas do tworzenia jednej, sztucznie złożonej funkcji celu (w której musielibyśmy z góry ustalić, co jest ważniejsze), pozwalają one wyznaczyć zbiór najlepszych dostępnych kompromisów. W praktyce oznacza to budowanie granicy rozwiązań

optymalnych, czyli linii lub płaszczyzny, na której znajdują się opcje, których nie da się już poprawić w jednym aspekcie bez jednoczesnego pogorszenia innego.

1.4. Reprezentacja rozwiązania, funkcja dopasowania i populacja

Każdy algorytm ewolucyjny wymaga trzech decyzji projektowych, bez których nie da się przejść od intuicji biologicznej do działającej procedury obliczeniowej. Po pierwsze trzeba ustalić reprezentację rozwiązania, a więc odpowiedzieć na pytanie, w jaki sposób zakodować parametry zadania w strukturze, na której będzie operować algorytm. Po drugie trzeba zdefiniować funkcję dopasowania, czyli miarę jakości rozwiązania. Po trzecie trzeba określić strukturę populacji i sposób inicjalizacji, bo to ona stanowi materiał wyjściowy dla procesu ewolucji.

Reprezentacja rozwiązania powinna być dostosowana do natury zadania. Dla parametrów ciągłych wygodne bywają wektory liczb rzeczywistych. Dla zadań kombinatorycznych, takich jak kolejność operacji czy trasa przejazdu, stosuje się permutacje albo specjalne kodowania pośrednie. W klasycznym algorytmie genetycznym często spotyka się kodowanie binarne, ponieważ ułatwia ono analizę, implementację oraz wprowadzenie podstawowych operatorów krzyżowania i mutacji. Należy jednak pamiętać, że kodowanie binarne nie jest warunkiem definicyjnym metody ewolucyjnej, lecz jedynie jedną z wielu możliwych reprezentacji.

Funkcja dopasowania pełni rolę środowiska selekcyjnego. To ona decyduje, które osobniki są uznawane za lepiej przystosowane, a które za słabsze. W prostych zadaniach dopasowanie może być bezpośrednio równe funkcji celu, na przykład wartości maksymalizowanej. W zadaniach inżynierskich bywa jednak znacznie bardziej złożone. Często trzeba uwzględnić kary za naruszenie ograniczeń, przeskalowanie różnych składników jakości, a czasem także uśrednianie wyników wielu symulacji, aby zmniejszyć wpływ szumu pomiarowego. Źle zaprojektowana funkcja dopasowania potrafi zniszczyć nawet dobrze dobrany algorytm, ponieważ kieruje poszukiwania w niewłaściwą stronę.

Populacja jest zbiorem kandydatów na rozwiązanie i zarazem pamięcią roboczą procesu ewolucji. Jej liczność wpływa na dwa przeciwstawne aspekty działania algorytmu. Zbyt mała populacja może bardzo szybko utracić różnorodność i skupić się na przeciętnym obszarze przestrzeni rozwiązań. Zbyt duża populacja zwiększa koszt obliczeń, co jest szczególnie dotkliwe wtedy, gdy ocena pojedynczego osobnika wymaga pełnej symulacji układu dynamicznego albo eksperymentu sprzętowego. W praktyce kompromis między licznością

populacji, kosztem obliczeń i poziomem różnorodności należy do podstawowych decyzji projektowych.

Przydatne jest także odróżnienie poziomu genotypu i fenotypu. Genotypem nazywa się strukturę zakodowaną, na przykład ciąg bitów lub wektor, na którym działają operatory genetyczne. Fenotyp to rzeczywiste rozwiązanie w przestrzeni problemu, a więc zestaw parametrów interpretowanych już w kontekście zadania. To rozróżnienie pomaga zrozumieć, że podobieństwo zapisów genotypowych nie zawsze musi oznaczać podobieństwo jakości rozwiązań fenotypowych, a zatem projekt reprezentacji nie może być traktowany jako decyzja czysto techniczna.

1.5. Podstawowe terminy biologiczne i ich odpowiedniki obliczeniowe

Tabela 1. Znaczenie poszczególnych terminów w algorytmach genetycznych/ewolucyjnych

Termin biologiczny	Znaczenie w algorytmach ewolucyjnych
chromosom	zakodowana reprezentacja rozwiązania
gen	elementarna składowa reprezentacji
allel	konkretna wartość genu
genotyp	struktura zapisu rozwiązania
fenotyp	rozwiązanie interpretowane w przestrzeni problemu
przystosowanie	wartość funkcji dopasowania

1.6. Operatory ewolucyjne: selekcja, krzyżowanie, mutacja i sukcesja

Mechanizm ewolucji w wersji obliczeniowej tworzą przede wszystkim operatory selekcji, rekombinacji, mutacji i sukcesji. Selekcja odpowiada za preferowanie lepiej przystosowanych osobników. Rekombinacja, zwana też krzyżowaniem, tworzy nowe rozwiązania przez łączenie informacji pochodzących od dwóch lub więcej rodziców. Mutacja wprowadza drobne zmiany losowe, które podtrzymują różnorodność i pozwalają odkrywać nowe rejony przestrzeni. Sukcesja określa natomiast, którzy osobnicy pozostają w populacji następnego pokolenia i jak silny ma być nacisk na elitaryzm, czyli zachowanie najlepszych dotychczas znalezionych rozwiązań.

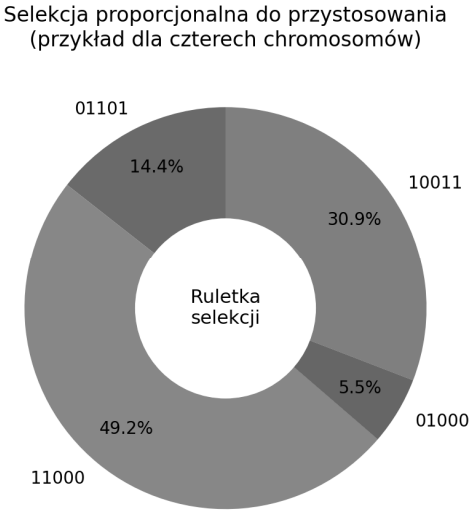
Najbardziej klasycznym przykładem selekcji jest selekcja proporcjonalna do przystosowania (rys. 3), znana obrazowo jako metoda ruletki. Każdemu osobnikowi przydziela

się udział proporcjonalny do wartości funkcji dopasowania, a następnie wielokrotnie losuje się rodziców. Rozwiązanie lepsze ma większe prawdopodobieństwo wyboru, ale słabsze osobniki nie są całkowicie eliminowane. Z dydaktycznego punktu widzenia metoda ruletki dobrze ilustruje zależność między jakością a szansą reprodukcji. W praktyce często stosuje się jednak również selekcję turniejową albo rankingową, ponieważ bywają bardziej stabilne numerycznie i mniej wrażliwe na skrajnie duże różnice w wartościach dopasowania.

Krzyżowanie jest operatorem, który najlepiej oddaje intuicję łączenia korzystnych cech dwóch rozwiązań. W najprostszym wariancie jednopunktowym wybiera się miejsce cięcia i zamienia końcowe fragmenty dwóch chromosomów (rys. 4). W reprezentacjach rzeczywistych stosuje się operatory arytmetyczne lub mieszające, a w permutacjach specjalne krzyżowania zachowujące poprawność kolejności. Skuteczność rekombinacji zależy od tego, czy w reprezentacji rzeczywiście istnieją użyteczne bloki informacji, które warto przekazywać potomstwu. W problemach, gdzie taka struktura nie jest zachowana, nadmierna wiara w krzyżowanie może prowadzić do pogorszenia jakości poszukiwań.

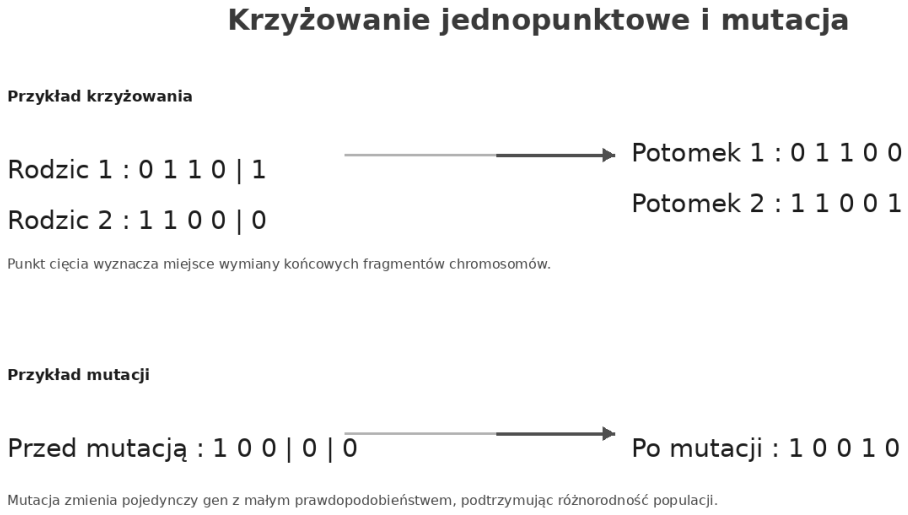
Mutacja ma zwykle niewielkie prawdopodobieństwo, lecz pełni rolę strategicznie ważną. To ona zapobiega zbyt szybkiemu ujednoliceniu populacji i utracie alleli, które mogłyby okazać się przydatne w dalszych etapach poszukiwań. W kodowaniu binarnym najczęściej oznacza zmianę bitu z zera na jeden lub odwrotnie (Rys. 4.). W reprezentacji rzeczywistej może przyjmować postać dodawania zaburzenia losowego, a w strategiach ewolucyjnych prowadzi do bardziej wyrafinowanej adaptacji skali mutacji. Zbyt słaba mutacja grozi przedwczesną zbieżnością, zbyt silna zamienia proces ewolucji w niemal losowe błądzenie. Z tego powodu operator ten powinien być strojony świadomie, a nie traktowany jako parametr pomocniczy bez większego znaczenia.

Sukcesja domyka pętlę ewolucyjną. W modelu generacyjnym cała populacja rodzicielska jest zastępowana nowym pokoleniem, natomiast w modelach stacjonarnych wymieniana jest tylko część osobników. Elitaryzm, czyli przeniesienie najlepszych rozwiązań do kolejnej generacji bez zmian, zwykle poprawia niezawodność algorytmu, bo zabezpiecza już osiągniętą jakość. Zbyt silny elitaryzm może jednak ograniczyć różnorodność, dlatego również w tym obszarze potrzebny jest kompromis między eksploatacją obiecujących rejonów a eksploracją nowych obszarów.



Rys. 3. Selekcja proporcjonalna do przystosowania na przykładzie czterech chromosomów

Źródło: Opracowanie własne.



Rys. 4. Krzyżowanie jednopunktowe i mutacja - przykłady operatorów podstawowych

Źródło: Opracowanie własne.

1.7. Klasyczny algorytm genetyczny

Klasyczny algorytm genetyczny można opisać jako powtarzaną sekwencję czynności: inicjalizacja populacji, ocena przystosowania, selekcja rodziców, krzyżowanie, mutacja, ocena potomstwa i decyzja o zatrzymaniu. Siła tej procedury nie polega na złożoności pojedynczego kroku, lecz na tym, że wszystkie kroki wzajemnie się uzupełniają. Selekcja wzmacnia

informację o jakości, krzyżowanie tworzy nowe kombinacje dotychczas znalezionych cech, a mutacja dostarcza kontrolowanej zmienności. W wyniku wielu iteracji populacja ma tendencję do przesuwania się w stronę rozwiązań coraz lepszych według przyjętego kryterium dopasowania.

Z punktu widzenia implementacji bardzo ważne jest rozróżnienie między strukturą algorytmu a jego konkretnymi ustawieniami. Sama struktura mówi, jakie kroki należy wykonać. Parametry mówią natomiast, jak intensywnie ma działać selekcja, jak często wykonywać krzyżowanie i mutację, jak liczna ma być populacja oraz jaki warunek stopu uznać za wystarczający. W klasycznych zastosowaniach często spotyka się prawdopodobieństwo krzyżowania z zakresu około 0,6-0,9 i prawdopodobieństwo mutacji wyraźnie mniejsze, ale nie istnieje jeden zestaw parametrów dobry dla wszystkich zadań. Dla problemów trudnych, zaszumionych albo wielokryterialnych wartości te mogą wymagać wyraźnie innych ustawień.

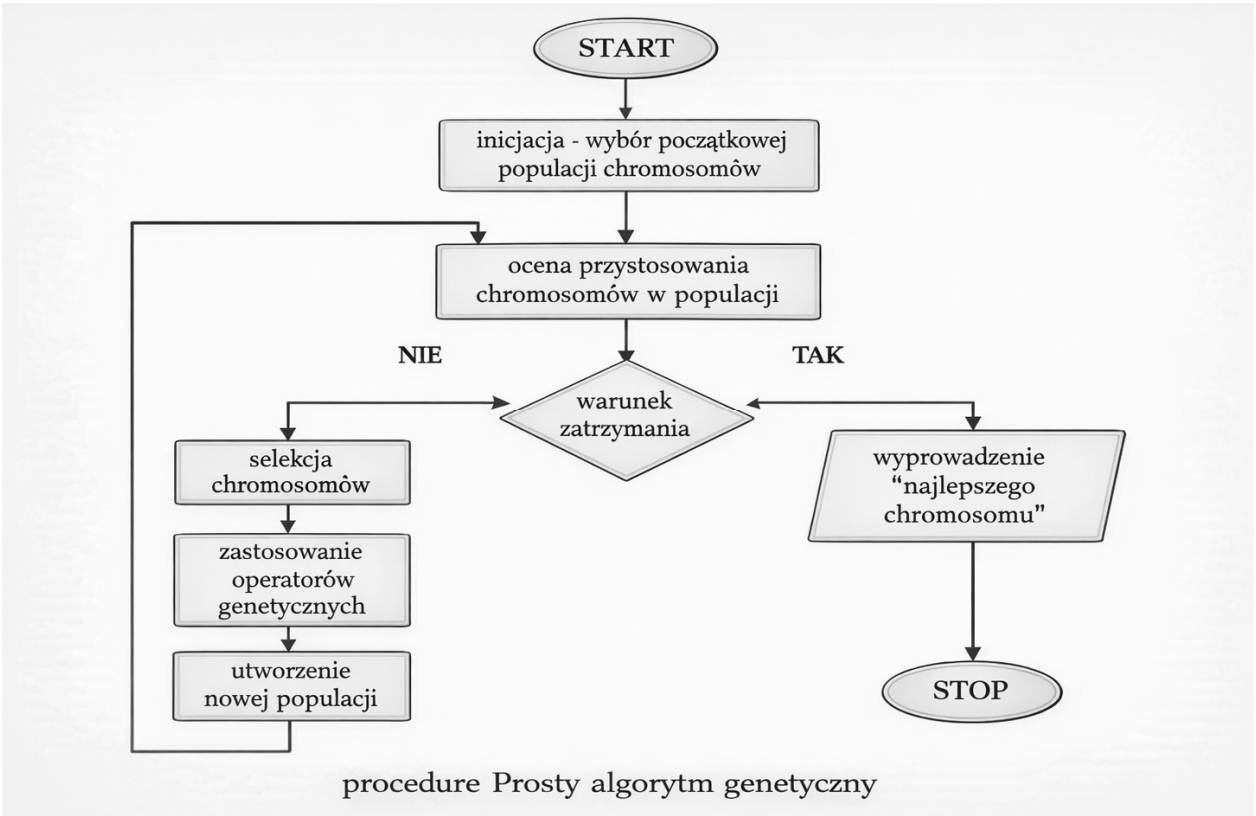
Warunek stopu powinien wynikać z celu obliczeń, a nie wyłącznie z przyzwyczajenia implementacyjnego. W praktyce zatrzymuje się algorytm po osiągnięciu określonej liczby pokoleń, po przekroczeniu budżetu obliczeniowego, po braku poprawy przez pewną liczbę iteracji albo po osiągnięciu rozwiązania spełniającego zadaną jakość. W zadaniach sterowania i robotyki ograniczenie budżetu czasowego ma znaczenie szczególne, ponieważ ocena rozwiązania może wymagać długiej symulacji lub eksperymentu na stanowisku, a zatem najbardziej sensownym miernikiem postępu bywa poprawa jakości w funkcji kosztu obliczeń, a nie tylko numer pokolenia.

Choć prosty algorytm genetyczny (AG) jest bardzo dobrym modelem dydaktycznym, nowoczesne implementacje często rozszerza się o mechanizmy dodatkowe: elitaryzm, ograniczanie duplikatów, adaptację parametrów, lokalne doskonalenie najlepszego osobnika albo hybrydyzację z metodami gradientowymi. W praktyce inżynierskiej takie połączenia bywają szczególnie efektywne. Algorytm ewolucyjny może wtedy pełnić rolę globalnego eksploratora przestrzeni rozwiązań, a metoda lokalna może dopracowywać znalezione obiecujące regiony. Poniżej został zapisany algorytmu genetycznego w postaci listy kroków – rys. 5.1, z kolei na rys. 5.2 w postaci schematu blokowego.

```
procedure Prosty algorytm genetyczny
begin
    t := 0
    inicjacja P^0
    ocena P^0
    while (not warunek stopu) do
    begin
        T^t := reprodukcja P^t
        O^t := krzyżowanie i mutacja T^t
        ocena O^t
        P^(t+1) := O^t
        t := t + 1
    end
end
```

Rys. 5.1. Schemat blokowy klasycznego algorytmu genetycznego

Źródło: Opracowanie własne.



Rys. 5.2. Schemat blokowy klasycznego algorytmu genetycznego

Źródło: Opracowanie własne.

1.8. Przykład działania algorytmu genetycznego

Aby uchwycić logikę działania klasycznego AG, warto prześledzić krótki przykład liczbowy na prostym zadaniu maksymalizacji funkcji $f(x) = x^2$ w przedziale $x \in [0,31]$. Zadanie jest celowo proste, ponieważ celem nie jest imponująca trudność obliczeniowa, lecz przejrzyste pokazanie procesu kodowania, oceny, selekcji i tworzenia potomstwa. W kodowaniu binarnym pięć bitów wystarcza do zapisania wszystkich liczb całkowitych z rozważanego zakresu.

Załóżmy populację początkową złożoną z czterech chromosomów: 01101, 11000, 01000 oraz 10011. Po dekodowaniu otrzymujemy odpowiednio liczby 13, 24, 8 i 19, a wartości funkcji dopasowania wynoszą odpowiednio169, 576, 64 i 361. Już na tym etapie widać, że osobnik reprezentujący $x = 24$ ma najlepsze dopasowanie, ale algorytm nie zatrzymuje się na prostym wyborze najlepszego chromosomu. Jego celem jest dalsze wykorzystanie informacji zawartej w populacji i wytworzenie nowych kandydatów.

Tabela 2. Początkowa populacja dla przykładu objaśniającego działanie AG

Nr	Chromosom	x	Przystosowanie $f(x) = x^2$
1	01101	13	169
2	11000	24	576
3	01000	8	64
4	10011	19	361

Jeżeli zastosujemy selekcję proporcjonalną do przystosowania, to chromosom o najwyższej wartości funkcji celu otrzyma największy udział w ruletce selekcji, ale pozostałe osobniki nie zostaną całkowicie wyeliminowane. Jest to ważne, ponieważ nawet rozwiązania słabsze mogą zawierać użyteczne fragmenty materiału genetycznego. Po selekcji tworzy się pulę rodzicielską, w której osobniki lepsze zwykle pojawiają się częściej niż gorsze. W naszym przykładzie najbardziej prawdopodobny jest wielokrotny wybór chromosomu 11000, natomiast chromosom 01000 ma najmniejszą szansę przetrwania.

Załóżmy następnie, że dwa wybrane chromosomy zostają skrzyżowane w jednym punkcie. Dla pary 01101 i 11000, po cięciu za czwartym bitem, powstają potomkowie 01100

oraz 11001, czyli odpowiednio liczby 12 i 25. Wartości dopasowania wynoszą wtedy 144 i 625. Jeden z potomków okazuje się lepszy niż najlepszy z rodziców, co dobrze ilustruje zasadniczą rolę rekombinacji: nie tylko kopiować, lecz także tworzyć nowe, lepsze kombinacje informacji.

Jeżeli dodatkowo dopuścimy mutację o małym prawdopodobieństwie, możliwa jest sporadyczna zmiana pojedynczego bitu. Taka zmiana sama w sobie nie gwarantuje poprawy, ale zabezpiecza populację przed utratą różnorodności. W dłuższym horyzoncie to właśnie równowaga między selekcją a zmiennością decyduje o tym, czy algorytm będzie systematycznie poprawiał jakość rozwiązań, czy też zbyt wcześnie ugrzęźnie w przeciętnym obszarze przestrzeni przeszukiwania.

Ten prosty przykład uczy trzech rzeczy. Po pierwsze, algorytm genetyczny nie działa jak losowe błądzenie, lecz wykorzystuje informację o jakości rozwiązań. Po drugie, poprawa średniego poziomu populacji jest zwykle ważniejsza niż jednorazowe znalezienie pojedynczego dobrego osobnika. Po trzecie, skuteczność całej procedury zależy od spójności reprezentacji, funkcji dopasowania i operatorów, a nie od samego faktu, że nazwano metodę algorytmem genetycznym.

1.9. Główne odmiany algorytmów ewolucyjnych

Po zrozumieniu klasycznego AG łatwo dostrzec, że rodzina algorytmów ewolucyjnych jest znacznie bogatsza. Strategie ewolucyjne szczególnie dobrze nadają się do zadań ciągłych i często operują na wektorach liczb rzeczywistych wraz z parametrami opisującymi skalę mutacji. Dzięki temu mogą dostosowywać intensywność przeszukiwania do kształtu problemu. W rozwiniętych wersjach, takich jak CMA-ES (ang. *Covariance Matrix Adaptation Evolution Strategy*), budowana jest nawet adaptacyjna informacja o strukturze kowariancji przestrzeni poszukiwań, co daje bardzo dobre wyniki w wielu zadaniach optymalizacji bezgradientowej.

Programowanie ewolucyjne historycznie koncentrowało się bardziej na mutacji i konkurencji rozwiązań niż na krzyżowaniu. Programowanie genetyczne rozszerza z kolei ideę ewolucji na struktury obliczeniowe, najczęściej drzewa reprezentujące programy, wyrażenia symboliczne albo reguły sterowania. Dla automatyki i robotyki jest to interesujące zwłaszcza wtedy, gdy nie optymalizuje się wyłącznie parametrów, lecz także samą strukturę prawa sterowania lub postać funkcji decyzyjnej.

Ewolucja różnicowa zajmuje szczególne miejsce w nowoczesnej praktyce inżynierskiej. Jej operator potomstwa powstaje przez dodawanie ważonych różnic między wektorami

rozwiązań, co prowadzi do zadziwiająco prostego, a zarazem skutecznego mechanizmu eksploracji przestrzeni ciągłej. W licznych zadaniach strojenia parametrów, identyfikacji modeli i optymalizacji wielowymiarowej DE (ang. *Differential Evolution*) bywa konkurencyjna lub lepsza od klasycznych GA (ang. *Genetic Algorithm*), zwłaszcza gdy problem ma naturalną reprezentację rzeczywistoliczbową.

Warto też wspomnieć o algorytmach wielokryterialnych, takich jak NSGA-II (ang. *Non-dominated Sorting Genetic Algorithm* II). Ich celem nie jest wyznaczenie jednego optimum, lecz zbioru rozwiązań niezdominowanych, tworzących granicę możliwości optymalizacyjnych układu. Dla projektowania układów automatyki i robotyki jest to podejście bardzo naturalne, bo pozwala analizować kompromisy między celami wzajemnie sprzecznymi. Inżynier może wówczas świadomie wybrać rozwiązanie preferujące na przykład oszczędność energii, szybkość działania albo dokładność pozycjonowania.

Tabela 3. Porównanie podstawowych klas algorytmów ewolucyjnych

Klasa	**Typ reprezentacji**	**Dominujący operator**	**Typowe zastosowania**	**Komentarz**
GA	ciągi symboli, bity, permutacje	selekcja + krzyżowanie	kombinatoryka, strojenie parametrów	klasyczny model dydaktyczny
ES	wektory liczb rzeczywistych	mutacja adaptacyjna	optymalizacja ciągła	silne powiązanie z samoadaptacją
EP	wektory / automaty	mutacja + konkurencja	modelowanie zachowania	mniejszy nacisk na rekombinację
GP	drzewa programów	rekombinacja poddrzew	synteza struktur i reguł	ewolucja programów i wyrażeń
DE	wektory liczb rzeczywistych	operator różnicowy	ciągła optymalizacja inżynierska	bardzo prosta, często bardzo skuteczna

Poniżej rozwinięcia skrótów z tabeli 3. Wszystkie techniki odnoszą się do głównych paradygmatów w dziedzinie Algorytmów Ewolucyjnych (ang. *Evolutionary Algorithms*):

- GA – Genetic Algorithm (Algorytm Genetyczny),

- ES – Evolution Strategy (Strategia Ewolucyjna),
- EP – Evolutionary Programming (Programowanie Ewolucyjne),
- GP – Genetic Programming (Programowanie Genetyczne),
- DE – Differential Evolution (Ewolucja Różnicowa).

1.10. Strojenie parametrów, ograniczenia i cele wielokryterialne

Strojenie parametrów algorytmu ewolucyjnego nie powinno być traktowane jako czynność drugorzędna. Liczność populacji, siła selekcji, prawdopodobieństwo lub skala mutacji, udział krzyżowania, rozmiar elity i warunek stopu współdecydują o dynamice całego procesu. W praktyce stosuje się dwie strategie. Pierwsza polega na dobraniu rozsądnych parametrów na podstawie literatury i kilku testów pilotażowych. Druga, bardziej zaawansowana, wykorzystuje adaptację parametrów w trakcie działania algorytmu, a więc przenosi część problemu strojenia do samego procesu optymalizacji.

Ograniczenia są integralną częścią większości zadań inżynierskich. Rozwiązanie może być bardzo dobre według funkcji celu, a jednocześnie całkowicie nieakceptowalne z punktu widzenia stabilności, bezpieczeństwa, zakresów pracy napędów albo geometrii ruchu. W metodach ewolucyjnych ograniczenia można obsługiwać na kilka sposobów: przez funkcje kar, przez naprawę niepoprawnych osobników, przez specjalne operatory zachowujące dopuszczalność albo przez selekcję preferującą najpierw rozwiązania dopuszczalne. Wybór metody zależy od problemu. Funkcje kar są najprostsze implementacyjnie, ale wymagają starannego skalowania, aby nie zdominować właściwej treści zadania.

W zadaniach wielokryterialnych szczególnie ważne jest unikanie zbyt wczesnej agregacji wszystkich celów do jednej liczby. Taka agregacja bywa wygodna, lecz często zaciera strukturę kompromisów. Populacyjny charakter algorytmów ewolucyjnych sprawia, że można równolegle utrzymywać wiele rozwiązań dobrych w różnych sensach. Daje to nie tylko bardziej informacyjny wynik, lecz także lepszy materiał do decyzji projektowej. W automatyce i robotyce, gdzie niemal zawsze występuje napięcie między szybkością, dokładnością, zużyciem energii i odpornością na zakłócenia, jest to zaleta o dużej wartości praktycznej.

1.11. Zastosowania w automatyce i robotyce

Zastosowania algorytmów ewolucyjnych w automatyce i robotyce są szerokie, ale warto patrzeć na nie przez pryzmat klas problemów, a nie listę modnych haseł. Pierwszą klasę stanowi

strojenie parametrów regulatorów i obserwatorów. Dotyczy to zarówno klasycznych regulatorów PID, jak i układów o większej liczbie parametrów, w których ręczne strojenie staje się czasochłonne i mało powtarzalne. Drugą klasą są zadania planowania trajektorii i ruchu, gdzie trzeba jednocześnie uwzględnić kinematykę, dynamikę, ograniczenia napędów oraz kryteria jakości ruchu.

Trzecią klasę tworzy identyfikacja i aproksymacja modeli. Algorytmy ewolucyjne mogą służyć do doboru parametrów modeli nieliniowych, selekcji cech, konfiguracji struktur rozmytych lub optymalizacji hiperparametrów innych metod uczenia. Czwartą klasą jest synteza struktur sterowania, a więc problemy, w których szuka się nie tylko wartości parametrów, lecz także samej postaci rozwiązania. W tym obszarze szczególnie interesujące są programowanie genetyczne i ewolucyjna robotyka, gdzie ewolucji może podlegać kontroler, architektura zachowania, a nawet częściowo morfologia układu.

Nie oznacza to oczywiście, że metody ewolucyjne są najlepsze w każdym przypadku. Jeżeli zadanie jest gładkie, dobrze uwarunkowane i posiada łatwo dostępne pochodne, klasyczne metody deterministyczne potrafią być szybsze i dokładniejsze. Siła metod ewolucyjnych ujawnia się przede wszystkim tam, gdzie model matematyczny jest trudny, kryteriów jest wiele, ograniczenia są złożone, a krajobraz funkcji celu nie sprzyja podejściom lokalnym. Innymi słowy, są to narzędzia szczególnie cenne wtedy, gdy rzeczywisty problem bardziej przypomina złożony projekt inżynierski niż eleganckie zadanie z analizy matematycznej.

1.12. Ograniczenia i dobre praktyki projektowe

Najczęstszym błędem początkującego użytkownika metod ewolucyjnych jest przekonanie, że sam wybór modnego algorytmu rozwiązuje problem projektowy. W rzeczywistości większość niepowodzeń wynika z trzech przyczyn: nietrafionej reprezentacji, źle zdefiniowanej funkcji dopasowania oraz niewłaściwego budżetu obliczeniowego. Jeżeli reprezentacja nie szanuje struktury problemu, krzyżowanie i mutacja będą niszczyły cenne informacje. Jeżeli funkcja dopasowania nie odzwierciedla rzeczywistej jakości rozwiązania, algorytm będzie zbiegał w stronę celu pozornego. Jeżeli zaś ocena pojedynczego osobnika jest zbyt kosztowna, to nawet dobrze zaprojektowana metoda okaże się niepraktyczna.

Drugim ograniczeniem jest brak gwarancji znalezienia optimum globalnego w skończonym czasie. Algorytmy ewolucyjne są metodami heurystycznymi i stochastycznymi.

Ich siła polega na odporności, elastyczności i szerokiej stosowalności, a nie na ścisłych gwarancjach optymalności dla dowolnego zadania. Z tego powodu dobre praktyki wymagają wykonywania wielu niezależnych uruchomień, raportowania rozrzutu wyników i porównywania metod na tym samym budżecie obliczeniowym. Pojedynczy udany eksperyment nie powinien być traktowany jako pełny dowód jakości algorytmu.

W projektach inżynierskich szczególnie opłacalne są podejścia hybrydowe. Metoda ewolucyjna może globalnie przeszukiwać przestrzeń rozwiązań, a po znalezieniu obiecującego rejonu można uruchomić procedurę lokalnego dostrajania. Warto także stosować walidację na modelach o rosnącej wierności, zaczynając od szybkich symulacji i przechodząc do bardziej kosztownych eksperymentów dopiero dla najlepszych kandydatów. Taki schemat obniża koszty obliczeniowe i zwiększa wiarygodność wyników.

Dobrą praktyką jest również jawne dokumentowanie całego procesu: sposobu kodowania, znaczenia parametrów, konstrukcji funkcji dopasowania, obsługi ograniczeń, warunku stopu oraz liczby niezależnych uruchomień. Bez takiej dokumentacji porównanie wyników jest niepełne, a implementacja trudna do odtworzenia. W dydaktyce automatyki i robotyki ma to znaczenie szczególne, ponieważ celem nie jest jedynie uruchomienie programu, lecz zrozumienie procesu projektowego i umiejętność krytycznej oceny otrzymanych rezultatów.

1.13. Podsumowanie

Algorytmy genetyczne i ewolucyjne należą do najważniejszych metod optymalizacji inspirowanych naturą. Ich wspólną cechą jest populacyjny charakter poszukiwań, wykorzystanie funkcji dopasowania oraz kontrolowane wprowadzanie zmienności. Algorytm genetyczny jest klasycznym i dydaktycznie bardzo wartościowym przedstawicielem tej rodziny, ale nie należy utożsamiać go z całością algorytmów ewolucyjnych. Z punktu widzenia zastosowań inżynierskich kluczowe jest nie tyle zapamiętanie nazw operatorów, ile zrozumienie zależności między reprezentacją rozwiązania, oceną jakości, różnorodnością populacji i strategią sukcesji.

Dla automatyki i robotyki metody ewolucyjne są szczególnie interesujące wtedy, gdy problem jest nieliniowy, wielokryterialny, ograniczony i trudny analitycznie. W takich warunkach zapewniają elastyczne, odporne i często bardzo praktyczne podejście do poszukiwania rozwiązań. Nie są jednak metodami magicznymi. Wymagają starannego projektowania i rzetelnej oceny eksperymentalnej. Dopiero wtedy stają się narzędziem, które

rzeczywiście wspiera inżyniera w budowie lepszych układów sterowania i bardziej autonomicznych systemów robotycznych.

2. Uczenie maszynowe

2.1. Wprowadzenie

Uczenie maszynowe stanowi dziś jeden z najważniejszych filarów współczesnej sztucznej inteligencji, ale jego znaczenie w inżynierii nie wynika z modnego słownictwa ani z marketingowej atrakcyjności samego pojęcia. Wynika ono z bardzo konkretnej potrzeby: coraz więcej systemów technicznych pracuje w środowisku bogatym w dane, złożonym, zmiennym i trudnym do opisania wyłącznie za pomocą klasycznych modeli analitycznych. W automatyce i robotyce obserwujemy to szczególnie wyraźnie. Nowoczesne linie technologiczne, układy mechatroniczne, roboty mobilne, systemy wizyjne i instalacje procesowe generują ogromne ilości informacji pochodzących z czujników, sterowników, rejestratorów zdarzeń i systemów nadzorczych. Sama dostępność danych nie rozwiązuje jednak żadnego problemu. Dopiero odpowiednio zbudowany model potrafi przekształcić surowy pomiar w przewidywanie, diagnozę, klasyfikację stanu, ocenę ryzyka albo decyzję sterującą. Właśnie w tym miejscu pojawia się uczenie maszynowe: jako zbiór metod, które pozwalają wydobywać z danych prawidłowości użyteczne z punktu widzenia projektowania i eksploatacji systemów technicznych.

W praktyce inżynierskiej uczenie maszynowe nie powinno być traktowane jako przeciwieństwo klasycznej teorii sterowania, identyfikacji obiektów czy analizy sygnałów. Jest raczej ich rozszerzeniem. Tam, gdzie model fizyczny jest dobrze znany, prosty i wystarczająco dokładny, rozwiązania analityczne pozostają najlepszym wyborem. Tam jednak, gdzie obiekt jest silnie nieliniowy, trudno obserwowalny, poddany licznym zakłóceniom albo zależny od wielu zmiennych jednocześnie, modelowanie oparte wyłącznie na równaniach bywa niewystarczające albo kosztowne. Uczenie maszynowe pozwala wtedy budować modele przybliżone, detektory stanów, klasyfikatory awarii, systemy predykcji jakości i algorytmy wspomagające sterowanie. Rozdział ten został napisany jako wprowadzenie dla studentów automatyki i robotyki. Jego celem nie jest zebranie maksymalnie dużej liczby nazw algorytmów, lecz uporządkowane objaśnienie logiki uczenia maszynowego: czym ono jest, jakie problemy rozwiązuje, jak buduje się modele uczące się, jak je oceniać oraz w jakich zadaniach technicznych wykorzystuje się je w sposób najbardziej sensowny i odpowiedzialny.

Rys. 6. Relacja między sztuczną inteligencją, uczeniem maszynowym i uczeniem głębokim

Źródło: Opracowanie własne.

2.2. Istota uczenia maszynowego

W literaturze klasycznej uczenie maszynowe definiuje się jako zdolność programu komputerowego do poprawy jakości wykonywania pewnego zadania na podstawie doświadczenia. Definicja ta, zaproponowana przez Toma M. Mitchella, pozostaje aktualna, ponieważ trafnie oddaje istotę zagadnienia: model nie otrzymuje pełnej, ręcznie zapisanej procedury postępowania, lecz uczy się zależności na podstawie danych. W ujęciu formalnym można powiedzieć, że zadanie polega na wyznaczeniu funkcji lub reguły decyzyjnej, która odwzorowuje dane wejściowe $\boldsymbol{x}$ w przewidywaną odpowiedź $\hat{\boldsymbol{y}}$. Odpowiedź ta może być liczbą rzeczywistą, etykietą klasy, prawdopodobieństwem, rankingiem, wykrytą strukturą danych albo sekwencją działań. Zależnie od rodzaju problemu model uczący się minimalizuje odpowiednio zdefiniowaną funkcję straty, czyli miarę błędu popełnianego względem danych uczących. Najprostszy zapis tej idei można przedstawić jako minimalizację empirycznego ryzyka:

$$\hat{R}(f) = (1/N) \Sigma L(y_i, f(x_i)),$$

gdzie $\boldsymbol{L}$ oznacza funkcję straty, a $\boldsymbol{f}$ jest modelem dobieranym na podstawie zbioru przykładów uczących. Chociaż zapis jest zwięzły, zawiera sedno całej dziedziny: z jednej strony mamy dane i sposób mierzenia błędu, z drugiej – rodzinę modeli, spośród których trzeba wybrać taki, który najlepiej uogólnia wiedzę na nowe przypadki.

Na tym etapie warto podkreślić dwie rzeczy, które bywają mylone w popularnych opisach. Po pierwsze, uczenie maszynowe nie polega na „rozumieniu" danych w ludzkim sensie. Model nie ma świadomości znaczenia pojęć, lecz operuje reprezentacją liczbową i optymalizuje parametry tak, aby poprawić wynik według zadanego kryterium. Po drugie, skuteczność modelu nie wynika jedynie z samego algorytmu. Ta sama metoda może działać bardzo dobrze albo bardzo słabo w zależności od jakości danych, sposobu przygotowania cech, doboru funkcji celu i procedury walidacji. Z punktu widzenia inżyniera ważniejsze od znajomości modnych nazw jest więc rozumienie pełnego procesu projektowania systemu uczącego się. Ostatecznym celem nie jest przecież uruchomienie algorytmu, lecz rozwiązanie konkretnego problemu technicznego w sposób powtarzalny, uzasadniony i bezpieczny.

Uczenie maszynowe należy odróżnić zarówno od całej sztucznej inteligencji, jak i od głębokiego uczenia. Sztuczna inteligencja jest pojęciem szerszym i obejmuje również systemy regułowe, planowanie, wnioskowanie symboliczne, wyszukiwanie heurystyczne czy algorytmy ewolucyjne. Uczenie maszynowe jest jedną z najważniejszych części tej dziedziny, lecz nie obejmuje jej w całości. Z kolei głębokie uczenie stanowi szczególny nurt uczenia maszynowego wykorzystujący wielowarstwowe modele neuronowe. W niniejszym rozdziale nacisk zostaje położony głównie na ogólne zasady uczenia maszynowego oraz klasyczne metody nadzorowane i nienadzorowane, ponieważ to one budują fundament pojęciowy potrzebny do świadomego studiowania kolejnych zagadnień, w tym sieci neuronowych i metod głębokich.

Współczesne ujęcie uczenia maszynowego jest silnie związane z teorią statystycznego uczenia. Zakłada ona, że dane obserwowane stanowią jedynie próbę z pewnego nieznanego rozkładu, a zadaniem modelu nie jest perfekcyjne odtworzenie tej próby, lecz możliwie dobre zachowanie na przyszłych danych pochodzących z tego samego lub podobnego procesu. Z tego powodu w praktyce minimalizuje się nie tylko błąd na zbiorze uczącym, ale często również człon regularyzacyjny ograniczający nadmierną złożoność modelu. Można to zapisać schematycznie jako minimalizację funkcji $\mathbf{J(f) = \hat{R}(f) + \lambda\Omega(f)}$, gdzie $\mathbf{\Omega(f)}$ opisuje złożoność modelu, a $\boldsymbol{\lambda}$ określa siłę regularyzacji. Ważna jest tutaj intuicja: model zbyt prosty nie opisze rzeczywistej zależności, a model zbyt złożony nauczy się nie tylko prawidłowości, lecz także przypadkowego szumu. Cała praktyka uczenia maszynowego jest w dużej mierze sztuką znajdowania właściwej równowagi między tymi dwoma skrajnościami.

2.3. Miejsce uczenia maszynowego w automatyce i robotyce

Znaczenie uczenia maszynowego w automatyce i robotyce bierze się z charakteru współczesnych układów technicznych. Obiekty przemysłowe, roboty współpracujące, systemy transportowe, układy predykcyjnego utrzymania ruchu, instalacje energetyczne czy stanowiska wizyjne pracują w warunkach, w których liczba dostępnych sygnałów jest bardzo duża, a zależności pomiędzy nimi bywają nieliniowe, zmienne w czasie i częściowo ukryte. W takich sytuacjach klasyczne modele fizyczne są nadal ważne, lecz często trzeba je wspomagać modelami opartymi na danych. Uczenie maszynowe umożliwia wówczas budowę tak zwanych miękkich czujników, czyli modeli szacujących wielkości trudne lub kosztowne do bezpośredniego pomiaru, a także klasyfikatorów stanów pracy, detektorów anomalii, modeli przewidujących jakość wyrobu i systemów wspierających decyzje operatora lub sterownika nadrzędnego.

W robotyce znaczenie uczenia maszynowego jest równie duże, ale ma nieco inną strukturę problemów. Tutaj obok klasycznej diagnostyki pojawia się percepcja, rozpoznawanie obiektów, lokalizacja, segmentacja sceny, estymacja ruchu, planowanie działań oraz uczenie zachowań w środowisku dynamicznym i częściowo niepewnym. Robot działający w rzeczywistym otoczeniu musi interpretować sygnały z kamer, lidarów, enkoderów, czujników siły, układów inercyjnych i szeregu innych źródeł. Oznacza to, że uczenie maszynowe staje się nie dodatkiem, lecz elementem architektury przetwarzania informacji. Jednocześnie systemy robotyczne są często układami bezpieczeństwa krytycznego. Z tego powodu metody uczące się nie mogą być oceniane wyłącznie przez pryzmat skuteczności na zbiorze testowym; muszą być także rozpatrywane pod kątem interpretowalności, odporności na zmianę warunków, stabilności, ograniczeń czasowych i możliwości bezpiecznej integracji z klasycznym sterowaniem. Właśnie dlatego w inżynierii tak duże znaczenie mają metody hybrydowe, łączące wiedzę fizyczną, modele matematyczne i algorytmy uczące się.

2.4. Podstawowe paradygmaty uczenia maszynowego

Najbardziej rozpowszechniony podział metod uczenia maszynowego obejmuje uczenie nadzorowane, uczenie nienadzorowane oraz uczenie ze wzmocnieniem. Każdy z tych paradygmatów odpowiada na nieco inny typ pytania, a zrozumienie tej różnicy jest ważniejsze niż pamiętanie długiej listy nazw algorytmów. W uczeniu nadzorowanym dysponujemy parami wejście–wyjście. Dla każdej próbki znamy wektor cech oraz prawidłową odpowiedź. Jeżeli odpowiedź jest liczbą rzeczywistą, mówimy o regresji; jeżeli należy do skończonego zbioru

klas, mówimy o klasyfikacji. Ten paradygmat dominuje w zastosowaniach przemysłowych, ponieważ wiele zadań daje się sformułować właśnie jako przewidywanie wielkości procesowej, rozpoznawanie stanu pracy maszyny, klasyfikowanie defektu lub estymowanie ryzyka awarii.

Uczenie nienadzorowane zakłada, że danych wejściowych jest dużo, ale nie są one opisane etykietami. Celem nie jest więc bezpośrednio przewidywanie zadanej odpowiedzi, lecz odkrycie struktury zbioru. W praktyce może to oznaczać grupowanie podobnych obserwacji, redukcję wymiaru, wykrywanie obserwacji odstających, identyfikację trybów pracy obiektu albo budowę bardziej zwartej reprezentacji danych. Dla inżyniera szczególnie ważne jest to, że wiele rzeczywistych systemów generuje ogromne ilości danych nieoznaczonych. Uczenie nienadzorowane pozwala w takich sytuacjach porządkować przestrzeń obserwacji i budować podstawę pod późniejszą diagnostykę, monitoring albo identyfikację nietypowych zachowań.

Uczenie ze wzmocnieniem ma jeszcze inną logikę. Nie opiera się na gotowych etykietach, lecz na interakcji agenta ze środowiskiem. Agent wykonuje działania, obserwuje ich skutki i otrzymuje sygnał nagrody, który mówi mu, czy dana sekwencja decyzji była korzystna. Celem jest znalezienie takiej polityki działania, która maksymalizuje skumulowaną nagrodę w czasie. Ten sposób uczenia jest szczególnie naturalny dla problemów sterowania sekwencyjnego, planowania ruchu i autonomicznego podejmowania decyzji. Jest jednak trudniejszy od klasycznej klasyfikacji czy regresji, ponieważ wymaga jednoczesnego rozwiązywania problemu eksploracji i eksploatacji, a w rzeczywistych układach technicznych dodatkowo musi respektować ograniczenia bezpieczeństwa, niezawodności i czasu reakcji.

W nowocześniejszym podziale coraz częściej wyróżnia się także uczenie półnadzorowane i samonadzorowane. Ma to znaczenie praktyczne, ponieważ w wielu zastosowaniach technicznych łatwo zgromadzić duże ilości danych nieoznaczonych, natomiast etykietowanie wymaga czasu eksperta i jest kosztowne. Uczenie półnadzorowane stara się wykorzystać niewielką liczbę przykładów opisanych oraz dużą liczbę przykładów nieopisanych. Uczenie samonadzorowane idzie krok dalej i buduje zadania pomocnicze wynikające z samej struktury danych, dzięki czemu model może nauczyć się użytecznej reprezentacji jeszcze przed etapem właściwego nadzoru. Chociaż metody te są bardziej rozpowszechnione w analizie obrazu, tekstu i sygnałów złożonych, ich ogólna idea jest cenna również dla automatyki: im trudniejsze i droższe jest ręczne etykietowanie danych, tym większą wartość mają metody potrafiące wykorzystać informację ukrytą w samych obserwacjach.

Rys. 7. Trzy podstawowe paradygmaty uczenia maszynowego

Źródło: Opracowanie własne.

Tabela 4. Porównanie podstawowych paradygmatów uczenia maszynowego

Paradygmat	**Dane uczące**	**Typowy wynik**	**Przykłady techniczne**
Nadzorowane	wejście + etykieta	klasa lub wartość liczbowa	diagnoza, predykcja jakości, prognozowanie
Nienadzorowane	same obserwacje	grupy, składowe, anomalia	monitoring procesu, analiza trybów pracy
Ze wzmocnieniem	interakcja i nagroda	polityka działania	sterowanie sekwencyjne, planowanie ruchu
Półnadzorowane / samonadzorowane	mało etykiet, dużo danych surowych	reprezentacja lub model pośredni	zadania z kosztownym etykietowaniem

2.5. Dane, cechy i reprezentacja problemu

W praktyce skuteczność systemu uczenia maszynowego zależy przede wszystkim od jakości reprezentacji problemu. Algorytm pracuje nie na „rzeczywistości", lecz na danych, a dane te są zawsze pewnym uproszczeniem rzeczywistego procesu. Dlatego pierwszy obowiązek projektanta polega na zrozumieniu, co właściwie oznacza pojedyncza próbka, jakie wielkości są obserwowane, z jaką częstotliwością są rejestrowane, jaki mają poziom szumu, czy są wzajemnie zsynchronizowane oraz które z nich rzeczywiście niosą informację istotną dla rozwiązania zadania. W automatyce i robotyce dane mogą mieć charakter tabelaryczny,

czasowy, obrazowy, akustyczny, przestrzenny albo wielomodalny. Każdy z tych typów wymaga innego spojrzenia na reprezentację.

Kluczowym pojęciem jest cecha, a więc taki opis próbki, który może zostać wykorzystany przez model. W prostych problemach cechami są bezpośrednie pomiary, na przykład temperatura, ciśnienie, prąd silnika, prędkość, położenie, liczba cykli pracy albo czas trwania określonego zdarzenia. W bardziej złożonych zadaniach cechy buduje się wtórnie. Dla sygnałów drganiowych mogą to być wartości skuteczne, kurtoza, współczynniki spektralne lub energia w wybranych pasmach częstotliwości. Dla obrazów będą to cechy geometryczne, teksturalne albo deskryptory wyuczane automatycznie przez model. W przypadku danych sekwencyjnych często trzeba uwzględniać opóźnienia, okna czasowe, pochodne numeryczne, zmienne skumulowane albo cechy opisujące dynamikę zmian. Cały ten etap określa się często mianem inżynierii cech i w klasycznym uczeniu maszynowym ma on znaczenie fundamentalne.

Równie ważne jak sam dobór cech jest przygotowanie danych. W praktyce trzeba zmierzyć się z brakami pomiarów, obserwacjami odstającymi, zmianą jednostek, różnymi zakresami wartości, nieciągłością rejestracji oraz ryzykiem przecieku informacji. Ten ostatni problem pojawia się wtedy, gdy w danych wejściowych znajdują się wielkości pośrednio ujawniające odpowiedź, którą model ma przewidywać, choć w warunkach rzeczywistych nie byłyby one dostępne. Błąd taki sztucznie zawyża wyniki eksperymentu i prowadzi do bardzo kosztownych rozczarowań po wdrożeniu. Dlatego przygotowanie danych nie jest etapem technicznym wykonywanym „na końcu”, lecz centralnym elementem całego projektu. W wielu zastosowaniach przemysłowych właśnie tu rozstrzyga się, czy model uczący się będzie użyteczny, czy pozostanie jedynie ciekawym eksperymentem.

W danych inżynierskich bardzo często konieczne są także operacje standaryzacji, normalizacji i skalowania. Jeżeli jedna cecha jest wyrażona w ułamkach milimetra, a inna w setkach amperów, to metody oparte na odległościach lub gradientach mogą faworyzować zmienną o większej skali, choć nie musi ona być bardziej informacyjna. Trzeba również zdecydować, jak postępować z brakami danych: czy usuwać niekompletne obserwacje, czy uzupełniać je estymacją, a może budować model odporny na brakujące wartości. W problemach czasowych pojawia się jeszcze kwestia poprawnego tworzenia okien obserwacyjnych i cech opóźnionych. Jeżeli model ma przewidywać stan układu w chwili $t + 1$, nie wolno wykorzystywać informacji, które w rzeczywistym użyciu byłyby dostępne dopiero później.

Tego rodzaju błędy są częste, bo pozornie poprawiają wyniki, ale faktycznie naruszają logikę zadania.

2.6. Typowy proces budowy modelu uczącego się

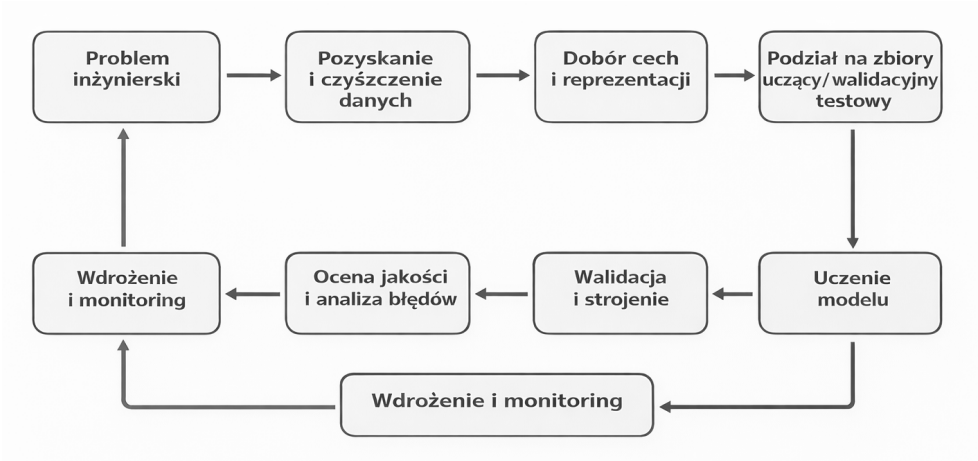
Budowa modelu uczenia maszynowego ma charakter iteracyjny, ale jej logika jest dość uporządkowana. Punktem wyjścia powinno być zawsze precyzyjne zdefiniowanie zadania. Trzeba ustalić, czy chodzi o klasyfikację, regresję, grupowanie, wykrywanie anomalii czy sterowanie sekwencyjne. Następnie należy określić, jak będzie mierzony sukces rozwiązania. W diagnostyce urządzeń ważniejsza może być wysoka czułość wykrywania uszkodzeń niż sama ogólna dokładność klasyfikacji. W predykcji jakości procesu bardziej istotne może być ograniczenie dużych odchyleń niż minimalizacja przeciętnego błędu. Już na tym etapie ujawnia się różnica między zadaniem matematycznym a rzeczywistym problemem inżynierskim: model ma być nie tylko „dobry statystycznie", lecz przede wszystkim użyteczny w danym kontekście eksploatacyjnym.

Po zdefiniowaniu celu następuje etap pozyskania i przygotowania danych, a następnie ich podział na zbiory uczący, walidacyjny i testowy. Zbiór uczący służy do dopasowania parametrów modelu. Zbiór walidacyjny pozwala dobrać hiperparametry, porównać konkurencyjne modele i kontrolować zjawisko przeuczenia. Zbiór testowy powinien zostać zachowany do końcowej, niezależnej oceny jakości. W danych czasowych i procesowych należy szczególnie uważać, aby podział nie niszczył naturalnej struktury obserwacji. Losowe mieszanie próbek może być właściwe w niektórych problemach tabelarycznych, ale w prognozowaniu lub diagnostyce sekwencyjnej prowadzi do nienaturalnie łatwego zadania i zawyżonych wyników.

Kolejny krok polega na wyborze modelu i procedury uczenia. W tym miejscu w praktyce popełnia się często dwa symetryczne błędy. Pierwszy polega na sięganiu po metodę najbardziej zaawansowaną obliczeniowo, choć prostszy model byłby równie skuteczny i łatwiejszy do interpretacji. Drugi polega na zbyt wczesnym przywiązaniu się do jednej techniki bez uczciwego porównania z sensownymi punktami odniesienia. Dobrą praktyką jest rozpoczęcie od modeli bazowych, takich jak regresja liniowa, regresja logistyczna, drzewo decyzyjne albo prosty klasyfikator odległościowy, a dopiero później przechodzenie do bardziej złożonych metod. Takie postępowanie pozwala lepiej zrozumieć naturę danych i uniknąć sytuacji, w której skomplikowany model maskuje źle postawione zadanie.

Ostatnim etapem jest wdrożenie i monitorowanie pracy modelu. W zastosowaniach przemysłowych jest to faza równie ważna jak samo uczenie. Model musi działać na danych napływających w czasie rzeczywistym albo w zadanym cyklu obliczeniowym, a jego wejścia i wyjścia muszą być poprawnie osadzone w architekturze systemu. Trzeba też obserwować, czy rozkład danych nie zmienia się w czasie, czy nie pogarsza się jakość przewidywań i czy nie pojawiają się nowe stany pracy, których model wcześniej nie widział. Oznacza to, że skuteczne uczenie maszynowe nie kończy się w chwili uzyskania dobrego wyniku eksperymentalnego. Prawdziwa wartość modelu ujawnia się dopiero wtedy, gdy potrafi on pracować stabilnie, powtarzalnie i odpowiedzialnie w rzeczywistym środowisku technicznym.

Na etapie strojenia modelu trzeba odróżnić parametry uczone automatycznie od hiperparametrów ustalanych przez projektanta. Parametry to na przykład współczynniki regresji albo podziały w drzewie decyzyjnym; hiperparametry określają z kolei strukturę lub ograniczenia modelu, takie jak głębokość drzewa, siła regularyzacji, liczba sąsiadów albo liczba składowych. Ich dobór odbywa się zwykle na podstawie walidacji. W profesjonalnej pracy inżynierskiej ważne jest również prowadzenie dziennika eksperymentów: zapisywanie wersji danych, sposobu przetwarzania, ustawień modelu i uzyskanych wyników. Bez tego nawet bardzo obiecujące rezultaty trudno odtworzyć, a brak powtarzalności jest jedną z najczęstszych przyczyn problemów przy przechodzeniu od etapu laboratoryjnego do wdrożenia.



Rys. 8. Iteracyjny cykl projektowania systemu uczenia maszynowego

Źródło: Opracowanie własne.

2.7. Klasyczne algorytmy uczenia nadzorowanego

W zadaniach nadzorowanych szczególne znaczenie mają algorytmy klasyczne, które mimo ogromnego rozwoju głębokiego uczenia wciąż pozostają podstawowym narzędziem w wielu zastosowaniach inżynierskich. Pierwszą grupę stanowią modele liniowe. Regresja liniowa, choć matematycznie prosta, jest niezwykle cenna jako punkt odniesienia i narzędzie interpretacji. Pozwala ocenić, które cechy są powiązane z przewidywaną zmienną, jaką mają siłę wpływu i czy problem rzeczywiście wymaga modelu nieliniowego. Wersje regularyzowane, takie jak ridge i lasso, dodatkowo pomagają ograniczać nadmierne dopasowanie i selekcjonować cechy. W klasyfikacji rolę analogiczną pełni regresja logistyczna, która modeluje prawdopodobieństwo przynależności do klasy i pozostaje jedną z najbardziej przejrzystych metod decyzyjnych.

Drugą ważną grupą są drzewa decyzyjne i modele zespołowe. Drzewo decyzyjne buduje hierarchię prostych pytań, na przykład „czy temperatura przekracza próg?” albo „czy energia sygnału w danym paśmie jest większa od wartości granicznej?”. Dzięki temu bywa łatwe do interpretacji, ale pojedyncze drzewo może mieć dużą wariancję i być wrażliwe na zmiany danych. Dlatego w praktyce często stosuje się lasy losowe oraz metody boostingowe. Lasy losowe łączą wiele drzew uczonych na różnych próbkach i podzbiorach cech, uzyskując większą stabilność oraz dobrą jakość dla danych tabelarycznych. Boosting buduje model etapami, koncentrując się na tych przykładach, które wcześniej były trudne do poprawnego przewidzenia. Metody te są obecnie bardzo silnym punktem odniesienia dla większości zadań klasyfikacji i regresji opartych na danych pomiarowych.

Kolejną rodziną są metody odległościowe i algorytmy oparte na marginesie decyzyjnym. Klasyfikator *k* najbliższych sąsiadów jest intuicyjny: nową obserwację przypisuje się do klasy reprezentowanej przez najbardziej podobne przykłady ze zbioru uczącego. Jego skuteczność zależy jednak silnie od doboru metryki, skali cech oraz liczby przykładów. Z kolei maszyny wektorów nośnych, czyli SVM, starają się znaleźć granicę decyzyjną maksymalizującą margines między klasami. Przy użyciu funkcji jądrowych mogą modelować także zależności nieliniowe. Przez wiele lat były jedną z najważniejszych metod w rozpoznawaniu wzorców i nadal są bardzo użyteczne wtedy, gdy liczba cech jest duża, a liczba przykładów umiarkowana.

Wybór metody nadzorowanej powinien wynikać z charakteru danych i wymagań aplikacji. Jeżeli istotna jest interpretowalność, pierwszeństwo mają modele liniowe, drzewa

i proste zespoły drzew. Jeżeli dane są niewielkie, ale dobrze opisane przez odpowiednie cechy, skuteczne mogą okazać się SVM albo modele odległościowe. Jeżeli problem ma charakter przemysłowy i dotyczy danych tabelarycznych, bardzo często najlepszy kompromis pomiędzy jakością, stabilnością i kosztem wdrożenia dają modele zespołowe. Najważniejsza zasada pozostaje jednak niezmienna: algorytm należy dobierać do zadania, a nie zadanie do algorytmu. W uczeniu maszynowym dojrzałość projektowa polega właśnie na tym, by nie utożsamiać wartości rozwiązania z poziomem jego złożoności.

Osobne miejsce zajmują modele probabilistyczne, w których przewidywanie ma postać rozkładu prawdopodobieństwa, a nie tylko pojedynczej decyzji. Z punktu widzenia eksploatacji systemu technicznego bywa to niezwykle cenne, ponieważ operator lub nadrzędny system sterujący może uwzględnić nie tylko sam wynik, ale również stopień pewności modelu. W prostszej postaci ideę tę reprezentuje na przykład naiwny klasyfikator Bayesa, a w bardziej rozbudowanej – różne odmiany modeli probabilistycznych i bayesowskich. Chociaż nie wszystkie z nich są tak popularne jak lasy losowe czy boosting, uczą ważnej zasady: model powinien nie tylko odpowiadać „co przewiduję?", ale także „jak bardzo jestem tego pewny?". W zastosowaniach bezpieczeństwa krytycznego taka informacja ma często równie duże znaczenie jak sama etykieta klasy lub wartość prognozy.

Tabela 5. Wybrane klasyczne algorytmy i ich typowe zastosowania

Rodzina metod	Mocne strony	Ograniczenia	Przykładowe użycie
Regresja liniowa / logistyczna	interpretowalność, szybkość, dobry punkt odniesienia	słabsze modelowanie silnych nieliniowości	predykcja wielkości procesowych, klasyfikacja stanów
Drzewa decyzyjne	łatwość objaśniania decyzji	pojedyncze drzewo może być niestabilne	reguły decyzyjne, szybka analiza danych tabelarycznych
Lasy losowe / boosting	wysoka skuteczność dla danych tabelarycznych	mniejsza przejrzystość niż model liniowy	diagnostyka, jakość procesu, scoring techniczny
SVM	dobry margines decyzyjny, skuteczność przy umiarkowanych zbiorach	wrażliwość na dobór parametrów i skalę cech	rozpoznawanie wzorców, klasyfikacja sygnałów
***n* najbliższych sąsiadów**	prostota i intuicyjność	koszt predykcji, zależność od metryki	zadania małej i średniej skali

2.8. Uczenie nienadzorowane i wykrywanie anomalii

Uczenie nienadzorowane jest szczególnie ważne tam, gdzie dostępne są duże zbiory danych, ale brak wiarygodnych etykiet. W takiej sytuacji celem staje się odkrycie struktury danych, a nie przewidywanie wcześniej zdefiniowanej odpowiedzi. Najbardziej znaną grupą metod są algorytmy grupowania. ***K***-średnich dzieli dane na zadaną liczbę skupień, minimalizując odległość obserwacji od środków klastrów. Jest metodą prostą i szybką, ale preferuje struktury o dość regularnym kształcie i wymaga wcześniejszego określenia liczby grup. Metody hierarchiczne budują natomiast drzewo podobieństwa, umożliwiając analizę danych na różnych poziomach szczegółowości. Algorytmy gęstościowe, takie jak DBSCAN, potrafią wykrywać skupienia o bardziej nieregularnej geometrii i jednocześnie wskazywać obserwacje odstające.

Drugim filarem uczenia nienadzorowanego jest redukcja wymiaru. W praktyce inżynierskiej liczba cech bywa duża, a część z nich jest wzajemnie skorelowana lub wnosi niewiele nowej informacji. Analiza głównych składowych, czyli PCA, pozwala opisać dane za pomocą nowego układu współrzędnych tak, aby możliwie duża część wariancji została zachowana przy mniejszej liczbie zmiennych. Metoda ta ma ogromne znaczenie dydaktyczne i praktyczne, ponieważ pokazuje, że bogaty opis procesu często można zastąpić kilkoma składowymi niosącymi najwięcej informacji. W monitorowaniu procesów i diagnostyce układów wielowymiarowych PCA jest narzędziem klasycznym, wciąż szeroko wykorzystywanym.

Szczególnym zastosowaniem uczenia nienadzorowanego jest wykrywanie anomalii. W wielu systemach przemysłowych dane reprezentujące poprawną pracę są liczne, natomiast awarie są rzadkie, niejednorodne i trudne do pełnego opisania. W takiej sytuacji sensowniejsze może być modelowanie zachowania normalnego niż uczenie klasyfikatora wszystkich możliwych uszkodzeń. Anomalię interpretuje się wtedy jako obserwację zbyt odległą od typowego wzorca pracy. Metody tego rodzaju wykorzystuje się w diagnostyce predykcyjnej, monitorowaniu sieci przemysłowych, analizie jakości procesu oraz w nadzorze ruchu robotów i manipulatorów. Ich zaletą jest mniejsza zależność od etykiet, ale wadą pozostaje trudność jednoznacznej interpretacji tego, czy odchylenie od normy oznacza rzeczywiście uszkodzenie, czy jedynie rzadki, lecz poprawny stan pracy.

W wykrywaniu anomalii stosuje się zarówno metody geometryczne, jak i probabilistyczne czy rekonstrukcyjne. Jednoklasowe maszyny wektorów nośnych próbują

wyznaczyć obszar opisujący normalne zachowanie układu, a obserwacje znajdujące się poza tym obszarem traktują jako podejrzane. Inne metody analizują lokalną gęstość danych albo błąd rekonstrukcji po przejściu przez model o ograniczonej pojemności. Z punktu widzenia automatyki najważniejsze jest jednak to, że anomalia nie jest pojęciem absolutnym. To, co w jednym trybie pracy jest zachowaniem poprawnym, w innym może oznaczać symptom uszkodzenia. Dlatego metody nienadzorowane muszą być projektowane z uwzględnieniem kontekstu procesowego, a ich wyniki powinny być interpretowane razem z wiedzą dziedzinową.

2.9. Uczenie ze wzmocnieniem

Uczenie ze wzmocnieniem zajmuje w uczeniu maszynowym miejsce szczególne, ponieważ zamiast przewidywać odpowiedź dla pojedynczej próbki, uczy się sekwencji decyzji. Agent obserwuje stan środowiska, wybiera działanie, a następnie otrzymuje nagrodę, która informuje go, czy decyzja była korzystna. Celem nie jest zatem maksymalizacja jednorazowej poprawności, lecz znalezienie polityki zapewniającej możliwie największy zysk w dłuższym horyzoncie czasu. Formalnie zadanie to wiąże się zwykle z procesami decyzyjnymi Markowa, ale z punktu widzenia studenta ważniejsza jest intuicja: agent musi nauczyć się działać tak, by krótkotrwałe korzyści nie prowadziły do gorszego wyniku w przyszłości. W sterowaniu, planowaniu ruchu i robotyce taka perspektywa jest naturalna, ponieważ każda decyzja zmienia stan układu i wpływa na możliwości kolejnych działań.

Zastosowania uczenia ze wzmocnieniem w automatyce i robotyce są bardzo obiecujące, lecz wymagające. Metoda ta dobrze pasuje do zadań sterowania adaptacyjnego, planowania trajektorii, manipulacji obiektami i uczenia złożonych zachowań ruchowych. Jednocześnie eksploracja niezbędna do uczenia może być niebezpieczna dla rzeczywistego obiektu technicznego. Robot nie może „uczyć się na błędach” w taki sam sposób jak agent w środowisku symulacyjnym, jeśli ceną błędu może być kolizja, uszkodzenie sprzętu albo zagrożenie dla człowieka. Dlatego współczesne podejścia kładą nacisk na uczenie bezpieczne, wykorzystanie modeli środowiska, filtrów bezpieczeństwa, uczenie w symulacji i transfer do rzeczywistego układu. Dla dydaktyki ważne jest więc nie tylko pokazanie potencjału uczenia ze wzmocnieniem, ale również wyraźne zaznaczenie jego ograniczeń i warunków odpowiedzialnego stosowania.

Jednym z centralnych problemów uczenia ze wzmocnieniem jest kompromis między eksploracją i eksploatacją. Agent powinien korzystać z działań, które już wydają się dobre, ale

jednocześnie musi od czasu do czasu sprawdzać inne możliwości, aby nie utknąć przy strategii tylko pozornie najlepszej. W robotyce i automatyce kompromis ten ma szczególnie trudny charakter, ponieważ eksploracja nie może prowadzić do zachowań niebezpiecznych. Dlatego duże znaczenie mają metody model-based, w których agent wykorzystuje model środowiska do planowania lub oceny skutków działań, oraz podejścia łączące uczenie z klasycznymi ograniczeniami sterowania. Współczesna literatura coraz częściej podkreśla, że sukces uczenia ze wzmocnieniem w zastosowaniach fizycznych zależy nie tylko od skutecznego algorytmu optymalizacji nagrody, ale przede wszystkim od poprawnego włączenia mechanizmów bezpieczeństwa i wiedzy o dynamice układu.

2.10. Uogólnianie, walidacja i ocena jakości modelu

Najważniejszą cechą dobrego modelu uczącego się nie jest to, że wiernie odtwarza dane, które już widział, lecz to, że poprawnie działa dla nowych przypadków. Właśnie tę zdolność nazywa się uogólnianiem. Jeśli model zbyt dokładnie dostosuje się do przypadkowych szczegółów i szumu obecnych w zbiorze uczącym, wówczas popełnia klasyczny błąd przeuczenia. Jeśli natomiast jest zbyt prosty, by uchwycić rzeczywistą strukturę zależności, mamy do czynienia z niedouczeniem. Między tymi skrajnościami istnieje kompromis, który w literaturze często opisuje się jako zależność bias-variance. Model o dużym obciążeniu jest zbyt sztywny i systematycznie popełnia błąd. Model o dużej wariancji nadmiernie reaguje na szczegóły danych uczących. Zadaniem projektanta jest znalezienie takiej złożoności modelu, przy której błąd na nowych danych będzie możliwie mały.

Ocena jakości modelu wymaga doboru właściwych miar. W klasyfikacji popularna jest dokładność, czyli udział poprawnie sklasyfikowanych przypadków, ale w wielu zastosowaniach technicznych jest to miara niewystarczająca. Jeżeli awarie są rzadkie, model może osiągać wysoką dokładność, a jednocześnie niemal nigdy ich nie wykrywać. Dlatego trzeba analizować także czułość, precyzję, specyficzność, miarę F1 oraz macierz pomyłek. W regresji typowe są błąd średniokwadratowy, pierwiastek z błędu średniokwadratowego, błąd bezwzględny i współczynnik determinacji. Wybór miary powinien odzwierciedlać rzeczywistą cenę błędu. Inaczej ocenia się model przewidujący temperaturę z dokładnością do ułamka stopnia, a inaczej klasyfikator mający sygnalizować stan awaryjny, w którym przeoczenie pojedynczego przypadku może być poważniejsze niż kilka fałszywych alarmów.

Walidacja jest procedurą, dzięki której próbujemy uczciwie oszacować zdolność modelu do uogólniania. Najczęściej polega na wydzieleniu zbioru walidacyjnego lub

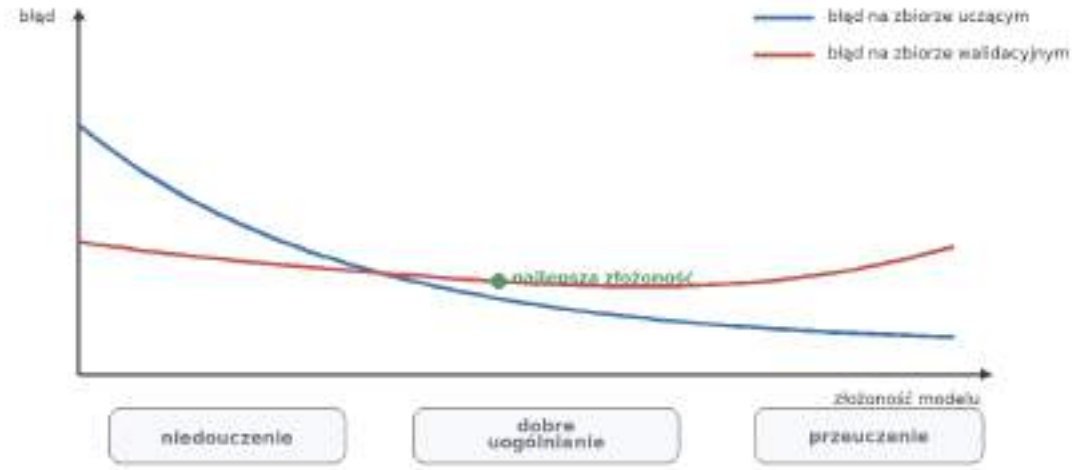
zastosowaniu walidacji krzyżowej. Jej sens jest prosty: model powinien być oceniany na danych, które nie brały udziału w jego bezpośrednim dopasowaniu. Tylko wtedy wynik ma wartość prognostyczną. W praktyce nie wolno wykorzystywać zbioru testowego do strojenia hiperparametrów, ponieważ prowadzi to do nieświadomego „uczenia się testu". Równie istotne jest respektowanie struktury czasowej i grupowej danych. Jeżeli próbki z tej samej serii pomiarowej albo z tego samego eksperymentu trafią równocześnie do zbioru uczącego i testowego, otrzymany wynik będzie zbyt optymistyczny.

Warto też pamiętać, że pojedyncza liczba opisująca jakość modelu nigdy nie daje pełnego obrazu. Model można oceniać globalnie, ale trzeba także patrzeć na profil błędów: dla jakich klas, zakresów pracy, typów zakłóceń i warunków eksploatacyjnych radzi sobie dobrze, a gdzie zawodzi. Taka analiza ma ogromne znaczenie w zastosowaniach przemysłowych, ponieważ system może być wdrażany właśnie tam, gdzie warunki są trudniejsze niż w danych uczących. Dojrzałe podejście do walidacji polega więc nie na szukaniu maksymalnie wysokiego wyniku, lecz na uczciwym oszacowaniu mocnych i słabych stron modelu oraz na ocenie, czy jego zachowanie pozostaje akceptowalne w rzeczywistym kontekście użytkowym.

Coraz większą rolę odgrywa również zagadnienie kalibracji i niepewności przewidywań. Dwa modele mogą mieć podobną dokładność klasyfikacji, ale różnić się tym, czy ich deklarowane prawdopodobieństwa są wiarygodne. Jeżeli model przypisuje klasie prawdopodobieństwo 0,95, użytkownik ma prawo oczekiwać, że takie przewidywania są rzeczywiście bardzo pewne. W zastosowaniach technicznych informacja o niepewności może decydować o tym, czy system działa autonomicznie, czy przekazuje sterowanie człowiekowi, czy uruchamia tryb bezpieczny, czy zleca dodatkowy pomiar. Oznacza to, że ocena modelu nie powinna kończyć się na prostym porównaniu średnich wartości miar jakości. Coraz częściej wymaga ona również analizy stabilności, kalibracji, odporności na zakłócenia oraz zachowania poza zakresem danych typowych.

Szczególnie ostrożnie należy traktować problemy niezrównoważenia klas. W diagnostyce technicznej i monitorowaniu awarii przypadki poprawnej pracy bywają tysiące razy liczniejsze niż przypadki uszkodzeń. W takiej sytuacji model może osiągać imponującą dokładność, po prostu przewidując klasę dominującą niemal zawsze. Dlatego analiza jakości musi uwzględniać nie tylko średni wynik globalny, ale również zdolność rozpoznawania klas rzadkich. Stosuje się wtedy odpowiednie miary, ważenie klas, dobór progów decyzyjnych, a niekiedy także metody zmiany rozkładu próbek w zbiorze uczącym. Dla inżyniera kluczowa

jest świadomość, że dobra miara jakości to taka, która odzwierciedla rzeczywiste ryzyko aplikacyjne, a nie tylko daje ładnie wyglądającą liczbę w raporcie.



Rys. 9. Zależność między złożonością modelu a błędem na zbiorze uczącym i walidacyjnym.

Źródło: Opracowanie własne.

2.11. Modele hybrydowe i znaczenie wiedzy dziedzinowej

Jednym z najważniejszych wniosków płynących ze współczesnej praktyki inżynierskiej jest to, że najwyższą wartość mają często nie modele całkowicie „czarne", ale rozwiązania hybrydowe, łączące wiedzę dziedzinową z uczeniem na danych. W automatyce i robotyce wiedza ta może przyjmować postać równań bilansu, ograniczeń fizycznych, warunków geometrycznych, modeli kinematycznych i dynamicznych, znanych praw sterowania albo logicznych relacji pomiędzy stanami procesu. Włączenie takich informacji do modelu uczącego się przynosi kilka korzyści jednocześnie. Po pierwsze, zmniejsza zapotrzebowanie na dane, ponieważ część struktury problemu jest już opisana przez wiedzę aprioryczną. Po drugie, poprawia interpretowalność modelu i ułatwia ocenę, czy jego decyzje są zgodne z intuicją inżynierską. Po trzecie, ogranicza ryzyko zachowań jawnie sprzecznych z fizyką obiektu, co w systemach bezpieczeństwa krytycznego ma znaczenie zasadnicze.

Przykładem takiego podejścia są miękkie czujniki oparte na połączeniu danych procesowych z uproszczonym modelem bilansowym, a także układy sterowania predykcyjnego, w których model uczący się pełni rolę członu korekcyjnego lub modelu zastępczego, ale decyzje sterujące nadal podlegają twardym ograniczeniom. W robotyce podobną rolę odgrywają metody łączące uczenie z kinematyką, modelem kontaktu, ograniczeniami siłowymi albo warstwą bezpieczeństwa nadzorującą plan ruchu. Rozwiązania

hybrydowe są często mniej efektowne z perspektywy demonstracji laboratoryjnej niż metody całkowicie oparte na danych, ale w zastosowaniach przemysłowych okazują się bardziej przewidywalne, łatwiejsze do uzasadnienia i lepiej przygotowane do pracy w warunkach rzeczywistych.

Wiedza dziedzinowa odgrywa także zasadniczą rolę w interpretacji wyników. Nawet bardzo skuteczny model może wskazywać zależności pozornie sensowne statystycznie, które z punktu widzenia procesu są przypadkowe albo wtórne. Inżynier powinien więc zawsze pytać, czy model wykorzystuje wielkości fizycznie wiarygodne, czy reaguje właściwie na zmiany warunków pracy oraz czy przewidywane trendy dają się pogodzić z wiedzą o obiekcie. Właśnie tu ujawnia się różnica między poprawnym uruchomieniem algorytmu a dojrzałym projektowaniem systemu uczącego się. Uczenie maszynowe nie eliminuje potrzeby rozumienia procesu; przeciwnie, czyni tę potrzebę jeszcze bardziej wyraźną, ponieważ błędnie postawione zadanie lub źle zinterpretowany wynik mogą prowadzić do bardzo przekonujących, lecz fałszywych wniosków.

2.12. Zastosowania uczenia maszynowego w automatyce i robotyce

Zakres zastosowań uczenia maszynowego w automatyce i robotyce jest bardzo szeroki, ale można wskazać kilka obszarów szczególnie reprezentatywnych. Pierwszy z nich to diagnostyka i utrzymanie predykcyjne. Na podstawie sygnałów drganiowych, akustycznych, prądowych, cieplnych i procesowych model może rozpoznawać stan łożysk, przekładni, silników, zaworów, pomp oraz narzędzi roboczych. Celem nie jest zwykle jedynie wykrycie już istniejącej awarii, lecz możliwie wczesne wskazanie symptomów pogarszania się stanu technicznego. Pozwala to lepiej planować przeglądy, ograniczać przestoje i przechodzić od utrzymania reaktywnego do predykcyjnego. Właśnie w tych zastosowaniach szczególnego znaczenia nabiera poprawne przygotowanie danych, ponieważ awarie są rzadkie, a warunki pracy obiektów różnią się między sobą.

Drugim ważnym obszarem jest modelowanie jakości i sterowanie procesami. W wielu instalacjach przemysłowych nie wszystkie wielkości są mierzalne w sposób ciągły albo tani. Uczenie maszynowe pozwala wtedy konstruować modele zastępcze, które na podstawie sygnałów pośrednich szacują jakość wyrobu, skład mieszaniny, stan procesu lub parametry trudne do bezpośredniego pomiaru. Takie miękkie czujniki są szeroko omawiane w literaturze procesowej i stanowią jedno z najbardziej dojrzałych zastosowań metod danych w inżynierii. Coraz częściej modele uczące się łączy się również z podejściem sterowania predykcyjnego,

budując układy hybrydowe, w których model danych wspomaga sterowanie przy zachowaniu ograniczeń i wymagań bezpieczeństwa.

Trzeci obszar to robotyka i systemy autonomiczne. Tutaj uczenie maszynowe wspiera percepcję, rozpoznawanie obiektów, lokalizację, estymację ruchu, segmentację sceny, planowanie i podejmowanie decyzji. W robotach mobilnych i manipulatorach bardzo duże znaczenie ma integracja danych z wielu sensorów oraz zdolność działania w zmiennym otoczeniu. Jednocześnie układy robotyczne silnie uwidaczniają problem tak zwanej luki między symulacją a rzeczywistością: model wyuczony w warunkach laboratoryjnych lub symulacyjnych nie zawsze zachowuje tę samą jakość po przeniesieniu do świata rzeczywistego. Dlatego współczesna robotyka coraz częściej poszukuje metod bezpiecznego uczenia, modeli hybrydowych oraz technik pozwalających łączyć dane z wiedzą o fizyce ruchu, ograniczeniach geometrycznych i dynamice układu.

Warto dodać, że uczenie maszynowe coraz wyraźniej wchodzi także do obszaru współpracy człowieka z systemem technicznym. W liniach zrobotyzowanych i stanowiskach montażowych modele uczące się mogą wspierać wykrywanie nieprawidłowych sekwencji działań, analizę ergonomii pracy, rozpoznawanie gestów, ocenę jakości interakcji człowiek – robot oraz adaptacyjne dostosowanie poziomu automatyzacji do sytuacji operacyjnej. W systemach wizyjnych stosuje się je do kontroli jakości, rozpoznawania części, śledzenia obiektów i inspekcji powierzchni. W tych zastosowaniach szczególnie widoczne jest to, że skuteczny model musi być częścią szerszego procesu inżynierskiego: obejmującego geometrię stanowiska, warunki oświetlenia, wymagania czasowe, sposób prezentacji wyników operatorowi i procedury postępowania w przypadku decyzji niepewnych.

2.13. Ograniczenia i dobre praktyki

Uczenie maszynowe bywa przedstawiane jako technologia uniwersalna, ale w praktyce ma ono wyraźne ograniczenia. Najważniejszym z nich jest zależność od danych. Model nie nauczy się poprawnych zależności, jeśli dane są ubogie, błędnie opisane, niezrównoważone lub zebrane w warunkach niereprezentatywnych dla przyszłej pracy systemu. Drugim ograniczeniem jest trudność interpretacji części metod, zwłaszcza bardziej złożonych modeli. W inżynierii problem ten ma znaczenie szczególne, ponieważ decyzja modelu często powinna być uzasadniona z punktu widzenia bezpieczeństwa, odpowiedzialności eksploatacyjnej i zaufania użytkownika. Trzecim ograniczeniem jest zmienność środowiska pracy. Nawet

dobrze oceniony model może z czasem tracić jakość wskutek starzenia aparatury, zmian konfiguracji obiektu, nowych trybów pracy czy przesunięcia rozkładu danych.

Z tych powodów dobre praktyki projektowe mają ogromne znaczenie. Po pierwsze, należy bardzo precyzyjnie zdefiniować zadanie i określić, jaki błąd jest naprawdę kosztowny w danej aplikacji. Po drugie, trzeba dokumentować pochodzenie danych, sposób ich czyszczenia, selekcji i podziału, aby zapewnić powtarzalność eksperymentu. Po trzecie, warto zaczynać od modeli prostych i interpretowalnych, a dopiero później zwiększać złożoność. Po czwarte, konieczna jest uczciwa walidacja na danych odseparowanych od procesu strojenia. Po piąte, po wdrożeniu model powinien być monitorowany tak samo jak inne elementy systemu technicznego. Nie można zakładać, że raz nauczony model będzie poprawnie działał w nieskończoność bez żadnej kontroli jakości.

Dojrzałe zastosowanie uczenia maszynowego w automatyce i robotyce wymaga więc nie tylko wiedzy algorytmicznej, ale również inżynierskiej dyscypliny. Najbardziej wartościowe rozwiązania to zwykle nie te, które wykorzystują najbardziej efektowną metodę, lecz te, które rozwiązują jasno zdefiniowany problem, są oparte na wiarygodnych danych, poddane rzetelnej walidacji i potrafią bezpiecznie współpracować z resztą systemu. W tym sensie uczenie maszynowe nie zwalnia inżyniera z odpowiedzialności za model; przeciwnie, wymaga od niego jeszcze większej świadomości założeń, ograniczeń i konsekwencji wdrożeniowych.

Do listy ograniczeń należy dodać również kwestie odpowiedzialności, cyberbezpieczeństwa i stronniczości danych. Model uczący się może utrwalać błędy obecne w danych historycznych, reagować nieprzewidywalnie na nietypowe wejścia albo być podatny na zakłócenia i celowe manipulacje. W środowiskach przemysłowych i autonomicznych oznacza to potrzebę projektowania architektury odpornej nie tylko statystycznie, lecz także systemowo. Dlatego coraz częściej mówi się o konieczności łączenia modeli uczących się z warstwą nadzorczą, która kontroluje dopuszczalność decyzji, ogranicza skutki błędów i zapewnia możliwość przejścia do trybu bezpiecznego. Dobra praktyka inżynierska zakłada ponadto, że nawet bardzo skuteczny model nie powinien być traktowany jako ostateczne źródło prawdy, lecz jako element systemu wspomagającego decyzję, którego działanie trzeba stale obserwować, testować i – w razie potrzeby – aktualizować.

2.14. Podsumowanie

Uczenie maszynowe jest dziś jedną z podstawowych metod analizy danych i budowy systemów inteligentnych, ale jego rzeczywista wartość ujawnia się dopiero wtedy, gdy zostaje poprawnie osadzone w kontekście zadania inżynierskiego. Dla studenta automatyki i robotyki najważniejsze jest zrozumienie, że model uczący się nie jest magicznym modułem, który samodzielnie „odkrywa prawdę", lecz narzędziem wymagającym świadomego zaprojektowania. Trzeba określić cel, przygotować dane, zbudować reprezentację problemu, dobrać model, ocenić jego zdolność uogólniania i rozważyć warunki wdrożenia. Dopiero taki ciąg działań prowadzi do rozwiązania, które ma znaczenie techniczne, a nie tylko akademickie.

Z perspektywy całego skryptu omawiany rozdział pełni rolę fundamentu pojęciowego. Wprowadza podstawowe języki opisu danych, modeli, błędów, walidacji i zastosowań, które będą potrzebne zarówno przy dalszym studiowaniu sztucznych sieci neuronowych i uczenia głębokiego, jak i przy analizie praktycznych zastosowań sztucznej inteligencji w automatyce i robotyce. Jeżeli czytelnik po lekturze rozumie różnicę między klasyfikacją a regresją, znaczenie cech i jakości danych, sens podziału na zbiory uczące i testowe, przyczyny przeuczenia oraz miejsce modeli uczących się w nowoczesnym systemie technicznym, to osiągnięty został cel najważniejszy: zbudowany został solidny punkt wyjścia do dalszego, bardziej zaawansowanego kształcenia.

3. Uczenie głębokie i sztuczne sieci neuronowe

3.1. Wprowadzenie

Sztuczne sieci neuronowe należą dziś do najważniejszych narzędzi współczesnej sztucznej inteligencji. Ich znaczenie nie wynika z tego, że stanowią modne rozszerzenie uczenia maszynowego, lecz z faktu, że pozwalają budować modele zdolne do aproksymowania złożonych, silnie nieliniowych zależności między danymi wejściowymi i wyjściowymi. Tam, gdzie klasyczne modele liniowe lub płytkie metody uczące się okazują się niewystarczające, sieci neuronowe potrafią stopniowo wydobywać z danych coraz bardziej abstrakcyjne reprezentacje, a następnie wykorzystywać je do klasyfikacji, regresji, segmentacji, sterowania czy predykcji sekwencji. Właśnie ta zdolność do reprezentacji wielopoziomowej leży u podstaw terminu „uczenie głębokie". Głębokość nie oznacza tutaj wyłącznie dużej liczby parametrów, ale przede wszystkim występowanie wielu kolejnych warstw przekształceń, z których każda buduje opis bardziej użyteczny dla rozwiązania danego zadania.

Z perspektywy automatyki i robotyki sieci neuronowe są szczególnie interesujące, ponieważ naturalnie wpisują się w problemy, w których sygnały pomiarowe są złożone, niepewne, zaszumione i wielowymiarowe. Dotyczy to rozpoznawania obiektów w obrazach z kamer przemysłowych, estymacji stanu z danych wieloczujnikowych, predykcyjnego utrzymania ruchu, identyfikacji uszkodzeń, analizy przebiegów czasowych, modelowania procesów technologicznych oraz planowania działań robota w zmiennym środowisku. W takich zadaniach nie wystarczy zwykle kilka ręcznie zdefiniowanych reguł. Potrzebny jest model, który potrafi nauczyć się reprezentacji danych na podstawie przykładów, a następnie uogólnić tę wiedzę na nowe przypadki. Uczenie głębokie nie eliminuje wiedzy inżynierskiej, lecz zmienia punkt ciężkości: część pracy, która dawniej polegała na ręcznym projektowaniu cech, dziś bywa przenoszona do architektury sieci i procesu uczenia.

Trzeba jednak od razu zaznaczyć, że pojęcia „sztuczne sieci neuronowe” i „uczenie głębokie” nie są dokładnie tym samym. Sztuczne sieci neuronowe tworzą szerszą klasę modeli inspirowanych uproszczoną ideą działania neuronów biologicznych. Uczenie głębokie oznacza z kolei ten obszar badań i zastosowań sieci neuronowych, w którym wykorzystuje się architektury wielowarstwowe uczące się reprezentacji hierarchicznych. Każde uczenie głębokie jest więc oparte na sieciach neuronowych, ale nie każda sieć neuronowa jest „głęboka” w sensie dzisiejszej praktyki. To rozróżnienie jest ważne dydaktycznie: pozwala zrozumieć, że nowoczesne głębokie modele wyrastają z klasycznych idei perceptronu, propagacji sygnału i korekcji wag, lecz rozwijają je do skali, która umożliwia pracę na obrazach, sygnałach czasowych, tekście czy danych multimodalnych.

3.2. Krótka geneza i znaczenie sieci neuronowych

Historia sztucznych sieci neuronowych jest dłuższa niż współczesny boom na uczenie głębokie. Już w połowie XX wieku rozwijano matematyczne modele uproszczonych neuronów i pierwsze koncepcje uczenia maszynowego opartego na adaptacji wag. W kolejnych dekadach pojawiały się zarówno okresy dużych nadziei, jak i okresy głębokiego sceptycyzmu, wynikające z ograniczeń obliczeniowych, niewielkich zbiorów danych oraz trudności w uczeniu bardziej złożonych architektur. Dopiero połączenie trzech czynników — wzrostu mocy obliczeniowej, dostępności dużych zbiorów danych i dopracowania metod optymalizacji — sprawiło, że sieci wielowarstwowe zaczęły uzyskiwać przewagę w wielu zadaniach praktycznych. Z tego powodu dzisiejsze uczenie głębokie należy rozumieć jako etap dojrzałości

całej rodziny modeli rozwijanych od wielu lat, a nie jako ideę, która pojawiła się nagle i bez wcześniejszego zaplecza teoretycznego.

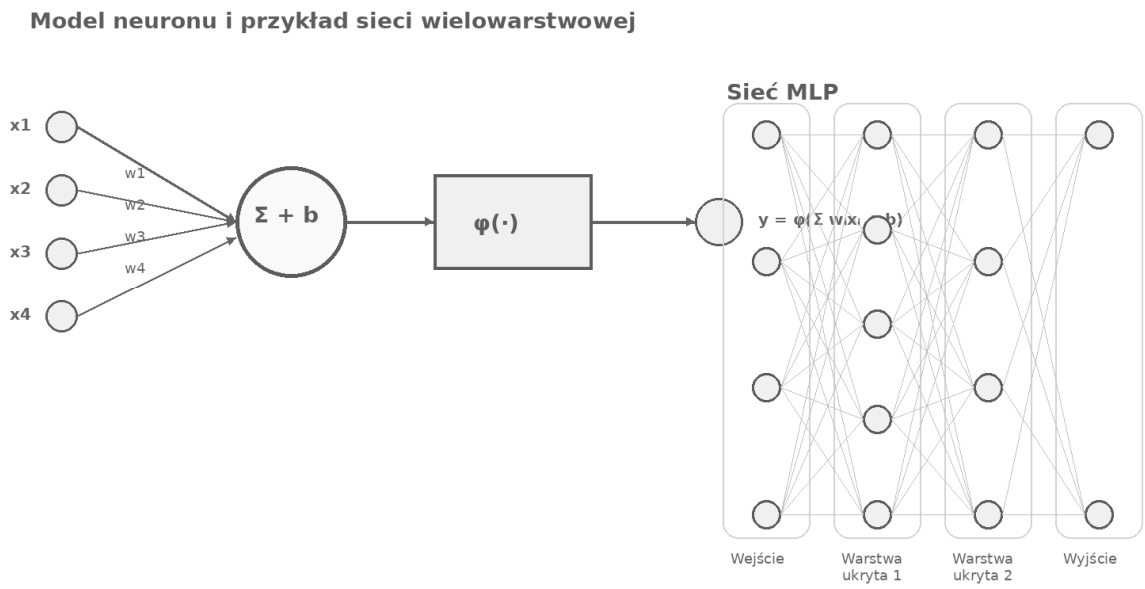
W literaturze przedmiotu za ważne punkty odniesienia uważa się perceptron, rozwój algorytmu propagacji wstecznej błędu, sukcesy sieci konwolucyjnych w rozpoznawaniu obrazów oraz późniejszy gwałtowny rozwój architektur rekurencyjnych, sekwencyjnych i opartych na mechanizmie uwagi. Współczesny inżynier powinien widzieć tę historię nie jako zbiór nazwisk i dat, ale jako stopniowe rozwiązywanie konkretnych problemów inżynierskich: jak skutecznie uczyć duże modele, jak ograniczać zanikający i eksplodujący gradient, jak wykorzystać lokalną strukturę obrazu, jak modelować zależności czasowe oraz jak radzić sobie z bardzo długimi sekwencjami. Każda kolejna architektura pojawiała się dlatego, że wcześniejsze modele miały praktyczne ograniczenia.

3.3. Model pojedynczego neuronu i perceptron wielowarstwowy

Najprostszym elementem sztucznej sieci neuronowej jest neuron formalny. Otrzymuje on na wejściu wartości liczbowe, mnoży je przez odpowiadające im wagi, sumuje wraz z wyrazem wolnym, a następnie podaje wynik na funkcję aktywacji. Zapis taki jest prosty, ale bardzo ważny, ponieważ pokazuje, że pojedynczy neuron jest w gruncie rzeczy nieliniowym przekształceniem wejść. Bez funkcji aktywacji sieć złożona nawet z wielu warstw byłaby jedynie złożeniem przekształceń liniowych, a więc w praktyce nadal odpowiadałaby pojedynczej transformacji liniowej. To właśnie nieliniowość umożliwia modelowanie złożonych granic decyzyjnych i relacji między zmiennymi.

W perceptronie wielowarstwowym, czyli klasycznym MLP, neurony są układane w warstwy. Warstwa wejściowa przekazuje dane do jednej lub wielu warstw ukrytych, a na końcu znajduje się warstwa wyjściowa odpowiedzialna za przewidywanie klasy, wartości liczbowej lub innego rodzaju odpowiedzi. Każda warstwa dokonuje własnego przekształcenia i przekazuje wynik dalej. Jeśli sieć ma wystarczającą liczbę neuronów i odpowiednio dobrane parametry, może aproksymować bardzo szeroką klasę funkcji. W praktyce nie wystarczy jednak sama „zdolność teoretyczna”; równie ważne są sposób uczenia, regularyzacja, dobór architektury i jakość danych. Należy zapamiętać, że sieć neuronowa nie działa dlatego, że naśladuje mózg w sensie biologicznej szczegółowości. Działa dlatego, że stanowi elastyczny model matematyczny, którego parametry mogą być dopasowywane na podstawie danych.

Dla zadań typowych dla automatyki i robotyki perceptron wielowarstwowy pozostaje ważnym punktem wyjścia. Gdy dane mają postać tabelaryczną lub zostały już opisane przez starannie zdefiniowany wektor cech, MLP potrafi dawać bardzo dobre wyniki przy umiarkowanej złożoności implementacyjnej. W dalszym ciągu jest też podstawą do zrozumienia bardziej złożonych architektur, ponieważ większość nowoczesnych sieci można traktować jako rozwinięcie tej samej idei: naprzemiennego stosowania transformacji liniowych, nieliniowości i mechanizmów sprzyjających stabilnemu uczeniu.



Rys. 10. Uproszczony model neuronu formalnego oraz perceptronu wielowarstwowego.

Źródło: Opracowanie własne.

3.4. Funkcje aktywacji i znaczenie nieliniowości

Funkcja aktywacji określa sposób, w jaki neuron reaguje na sumę ważoną swoich wejść. Historycznie stosowano funkcje progowe oraz sigmoidalne, które miały bezpośrednią interpretację związaną z „pobudzeniem" neuronu. W praktyce współczesnego głębokiego uczenia najczęściej wykorzystuje się jednak funkcje, które sprzyjają stabilnej optymalizacji i nie powodują nadmiernego zaniku gradientu. Klasycznym przykładem jest funkcja ReLU, czyli Rectified Linear Unit, która dla dodatnich argumentów zachowuje się liniowo, a dla

ujemnych zwraca zero. Jej popularność wynika z prostoty, małego kosztu obliczeniowego i korzystnych własności numerycznych.

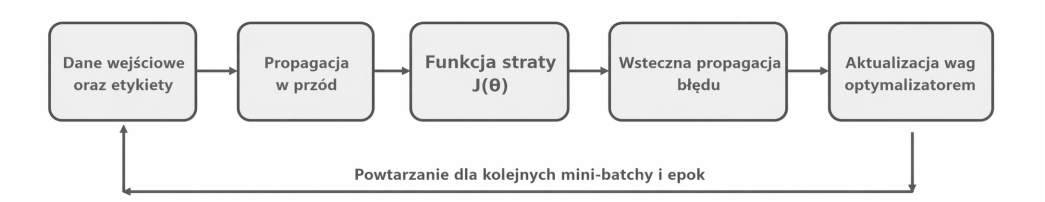
W zależności od zadania stosuje się również odmiany ReLU, funkcję tanh, sigmoidalną, a w warstwach wyjściowych — funkcje dobrane do znaczenia predykcji, takie jak softmax dla klasyfikacji wieloklasowej czy funkcja liniowa dla regresji. Nie jest to detal techniczny. Wybór funkcji aktywacji wpływa na dynamikę uczenia, zakres odpowiedzi modelu, interpretację wyjścia oraz łatwość optymalizacji. W sensie praktycznym można powiedzieć, że funkcje aktywacji są jednym z mechanizmów nadających sieci zdolność do opisu nieliniowych własności procesu. Bez nich głębokość architektury traciłaby znaczną część sensu.

3.5. Jak sieć się uczy: funkcja straty, gradient i propagacja wsteczna

Uczenie sieci neuronowej polega na takim doborze wag i wyrazów wolnych, aby zmniejszyć wartość funkcji straty opisującej rozbieżność między przewidywaniami modelu a poprawnymi odpowiedziami. Jeżeli zadaniem jest regresja, funkcja straty może opierać się na błędzie średniokwadratowym lub błędzie bezwzględnym. Jeśli zadaniem jest klasyfikacja, zwykle stosuje się warianty entropii krzyżowej. W praktyce funkcja straty jest matematycznym streszczeniem celu, jaki stawiamy modelowi. To ona mówi, co znaczy „uczyć się dobrze” w danym problemie. Dlatego nie można jej wybierać przypadkowo; musi wynikać z natury zadania i wymagań aplikacyjnych.

Najważniejszym algorytmem uczenia sieci wielowarstwowych jest propagacja wsteczna błędu, czyli backpropagation. Jej idea polega na wykorzystaniu reguły łańcuchowej rachunku różniczkowego do obliczenia, jak zmiana każdego parametru wpływa na wartość funkcji straty. Dzięki temu możliwe staje się wyznaczenie gradientu względem wszystkich wag w czasie obliczeniowo akceptowalnym nawet dla dużych modeli. Następnie parametry są aktualizowane zgodnie z wybraną metodą optymalizacji, najczęściej opartą na spadku gradientowym i jego wariantach mini-batchowych. Intuicyjnie można powiedzieć, że model najpierw wykonuje przejście w przód, obliczając wynik, a potem „cofa” informację

o popełnionym błędzie przez kolejne warstwy, ustalając, które parametry należy skorygować i w jakim kierunku.



Rys. 11. Cykl uczenia sieci neuronowej: dane wejściowe, przejście w przód, obliczenie straty, propagacja wsteczna i aktualizacja parametrów

Źródło: Opracowanie własne.

Z metodycznego punktu widzenia szczególnie ważne jest, aby nie traktować propagacji wstecznej jako magicznego rytuału składającego się z wzorów do zapamiętania. Jest to konsekwentna procedura obliczania pochodnych w złożonym układzie funkcji. Dopiero na tym fundamencie buduje się nowoczesne praktyki uczenia: mini-batche, harmonogramy współczynnika uczenia, momentum, adaptacyjne metody optymalizacji i strategie zatrzymywania treningu. Student, który zrozumie logikę przejścia w przód, funkcji straty i przekazywania gradientu, będzie znacznie łatwiej interpretował zachowanie dowolnej architektury głębokiej.

3.6. Optymalizacja, inicjalizacja i stabilność uczenia

W praktyce uczenie głębokich sieci neuronowych jest znacznie bardziej delikatne niż uczenie prostych modeli liniowych. Parametrów jest dużo, funkcja celu bywa bardzo złożona, a krajobraz optymalizacji nie ma postaci pojedynczej gładkiej misy. Z tego powodu duże znaczenie mają pozornie techniczne elementy, takie jak inicjalizacja wag, dobór współczynnika uczenia, rozmiar partii danych, czy użycie algorytmów optymalizacyjnych typu SGD z momentum, RMSProp lub Adam. Zbyt duży krok uczenia może prowadzić do niestabilności, a zbyt mały do powolnej zbieżności lub utknięcia w słabych rozwiązaniach. Również zła inicjalizacja może sprawić, że sygnał i gradient będą stopniowo zanikać lub gwałtownie rosnąć.

Jednym z podstawowych problemów w głębokich architekturach był historycznie zanikający i eksplodujący gradient. Jeśli kolejne pochodne mają wartości bardzo małe, informacja o błędzie słabnie w miarę cofania się przez warstwy, przez co wcześniejsze części

sieci uczą się zbyt wolno. Jeśli natomiast pochodne gwałtownie rosną, trening staje się niestabilny. Rozwiązania tego problemu obejmują odpowiedni dobór funkcji aktywacji, inicjalizacji parametrów, architektur rezydualnych, normalizacji i technik przycinania gradientu. Warto podkreślić, że sukces nowoczesnego uczenia głębokiego wynika nie tylko z „większych sieci", ale właśnie z rozwiązania szeregu praktycznych problemów numerycznych i optymalizacyjnych.

Z perspektywy zastosowań inżynierskich oznacza to prostą, ale ważną zasadę: jakość modelu nie zależy wyłącznie od jego teorii reprezentacyjnej. Zależy również od tego, czy proces uczenia został zaprojektowany tak, aby wykorzystać tę zdolność reprezentacji. Model o dużym potencjale może okazać się bezwartościowy, jeśli uczenie jest niestabilne, dane są źle przygotowane, a hiperparametry dobrane przypadkowo.

3.7. Generalizacja, przeuczenie i regularyzacja

Podobnie jak w całym uczeniu maszynowym, prawdziwym testem jakości sieci neuronowej nie jest to, jak dobrze zapamiętała przykłady uczące, lecz to, jak radzi sobie z danymi, których wcześniej nie widziała. Problem przeuczenia pojawia się wtedy, gdy model nadmiernie dopasowuje się do szczegółów zbioru treningowego, w tym do szumu i przypadkowych fluktuacji, tracąc zdolność uogólniania. Ryzyko to jest szczególnie duże w głębokich sieciach, ponieważ ich pojemność modelowa jest bardzo wysoka. Jeśli liczba parametrów jest duża, a danych stosunkowo mało, model może uzyskać świetne wyniki na zbiorze uczącym i jednocześnie zawodzić po wdrożeniu.

W praktyce stosuje się kilka głównych klas środków zaradczych. Należą do nich odpowiedni podział danych na zbiory uczące, walidacyjne i testowe, wczesne zatrzymywanie uczenia, regularyzacja wag, dropout, augmentacja danych oraz normalizacja. Szczególną rolę odgrywa augmentacja, zwłaszcza w zadaniach wizyjnych i akustycznych, ponieważ pozwala sztucznie zwiększyć różnorodność zbioru uczącego bez konieczności kosztownego zbierania nowych danych. W inżynierii produkcyjnej bywa to niezwykle istotne: wadliwe produkty, rzadkie awarie czy nietypowe warunki pracy nie zawsze są reprezentowane przez dużą liczbę próbek, a model musi umieć radzić sobie właśnie z nimi.

Do regularyzacji warto podchodzić nie jak do pojedynczego triku, lecz jak do ogólnej filozofii projektowania modelu. Obejmuje ona rozsądny dobór wielkości sieci, dyscyplinę

walidacji, kontrolę rozkładu danych, analizę błędów oraz stopniowe zwiększanie złożoności architektury. Często lepsze wyniki daje architektura umiarkowanie złożona, ale poprawnie przygotowana i dobrze zwalidowana, niż bardzo duża sieć trenowana bez planu i bez zrozumienia ograniczeń danych.

3.8. Normalizacja i przygotowanie danych

W głębokim uczeniu jakość danych bywa równie ważna jak sama architektura. Dane muszą być poprawnie zsynchronizowane, oczyszczone i reprezentatywne względem rzeczywistych warunków pracy. W automatyce i robotyce część trudności wynika z tego, że sygnały pochodzą z wielu czujników, mają różne częstotliwości próbkowania, bywają obarczone opóźnieniami, brakami i dryfem aparatury pomiarowej. Jeśli te problemy zostaną zignorowane, nawet bardzo dobra sieć nie będzie w stanie nauczyć się wiarygodnego modelu. Właśnie dlatego etap przygotowania danych powinien być traktowany jako integralna część projektu, a nie jako czynność pomocnicza wykonywana przed „właściwym" uczeniem.

Ważne znaczenie ma także normalizacja wejść. Zmienne o bardzo różnych skalach mogą prowadzić do trudności optymalizacyjnych, dlatego często stosuje się standaryzację lub przeskalowanie do ustalonego zakresu. Wewnątrz sieci podobną funkcję pełnią techniki normalizacyjne, z których najbardziej znana jest batch normalization. Ich rola polega między innymi na stabilizacji rozkładu aktywacji i ułatwieniu uczenia głębokich architektur. W praktyce jednak nie ma jednej uniwersalnej procedury. Dane obrazowe, dane czasowe, chmury punktów, przebiegi wibracyjne i wielokanałowe sygnały procesowe wymagają odmiennych decyzji dotyczących okien czasowych, filtrowania, deskryptorów i sposobu etykietowania.

3.9. Konwolucyjne sieci neuronowe i ich znaczenie w wizji przemysłowej

Sieci konwolucyjne, czyli CNN, stanowią najważniejszą klasę architektur dla zadań wizyjnych i dla wielu innych problemów, w których istotne znaczenie ma lokalna struktura danych. Podstawową ideą CNN jest stosowanie filtrów konwolucyjnych przesuwanych po obrazie lub sygnale. Filtry te uczą się lokalnych wzorców, takich jak krawędzie, tekstury, proste kształty, a w głębszych warstwach także bardziej złożone obiekty i relacje przestrzenne. Kluczową zaletą jest współdzielenie wag: ten sam filtr działa w wielu miejscach obrazu. Dzięki temu liczba parametrów jest znacznie mniejsza niż w klasycznych w pełni połączonych

warstwach, co pozwala skuteczniej wykorzystywać dane obrazowe i zmniejsza ryzyko przeuczenia.

W automatyce i robotyce znaczenie CNN jest ogromne. W inspekcji wizyjnej wykrywają one wady powierzchni, pęknięcia, niezgodności montażowe i defekty geometryczne. W robotyce mobilnej pomagają rozpoznawać obiekty, segmentować scenę, estymować położenie elementów oraz wspierać nawigację. W manipulatorach przemysłowych mogą uczestniczyć w lokalizacji części, ocenie jakości chwytu i kontroli orientacji obiektu. W praktyce przemysłowej ogromne znaczenie ma to, że sieć konwolucyjna może uczyć się bezpośrednio z obrazów surowych albo z map cech wygenerowanych przez wcześniejsze etapy systemu. Nie oznacza to jednak, że CNN zawsze można stosować bezrefleksyjnie. Obrazy przemysłowe są często znacznie trudniejsze niż klasyczne obrazy konsumenckie: tło bywa jednorodne, różnice między klasami subtelne, a liczba przykładów mała. Do tego dochodzą zmiany oświetlenia, odbicia, zanieczyszczenia optyki i wahania geometrii obserwacji. Dlatego skuteczność CNN w zastosowaniach inżynierskich silnie zależy od jakości akwizycji obrazu, procedur augmentacji oraz przemyślanego doboru architektury i metryki oceny.

Nowoczesne systemy wizyjne rzadko kończą się na prostym klasyfikatorze. W praktyce spotyka się architektury detekcyjne, segmentacyjne i hybrydowe, a także sieci rezydualne, które ułatwiają uczenie bardzo głębokich modeli. Warto zapamiętać, że sukces CNN wynika nie tylko z samej konwolucji, ale z całego ekosystemu rozwiązań: poolingów, połączeń skrótowych, normalizacji, augmentacji i transfer learningu.

3.10. Modele sekwencyjne: RNN, LSTM i GRU

Dla danych czasowych i sekwencyjnych naturalnym punktem wyjścia były przez długi czas sieci rekurencyjne. Ich najważniejszą cechą jest obecność stanu ukrytego, który przenosi informację pomiędzy kolejnymi krokami sekwencji. Dzięki temu model nie analizuje każdej próbki w całkowitym oderwaniu od przeszłości, lecz może uwzględniać kontekst czasowy. W zastosowaniach automatyki i robotyki takie właściwości są wyjątkowo ważne: przebiegi z czujników, trajektorie ruchu, logi procesu czy sekwencje poleceń mają charakter uporządkowany w czasie i nie można interpretować ich jako przypadkowych zbiorów punktów.

Klasyczne RNN okazały się jednak trudne do uczenia przy dłuższych zależnościach czasowych. Powodem był głównie zanikający lub eksplodujący gradient. Rozwiązaniem stały

się architektury LSTM i GRU, które wyposażono w mechanizmy bramek kontrolujących zapamiętywanie, zapominanie i aktualizację informacji. Dzięki temu model może zdecydować, które elementy historii są ważne, a które należy odrzucić. W praktyce LSTM i GRU długo stanowiły podstawowe narzędzie w zadaniach takich jak prognozowanie sygnałów, analiza drgań, wykrywanie zdarzeń w czasie, rozpoznawanie sekwencji działań robota czy modelowanie procesów dynamicznych.

Z punktu widzenia dydaktyki istotne jest zrozumienie, że modele sekwencyjne nie rozwiązują automatycznie wszystkich problemów związanych z czasem. Trzeba jeszcze określić długość okna obserwacji, sposób kodowania sekwencji, częstotliwość próbkowania, synchronizację danych z wielu źródeł oraz sensowną metrykę oceny. W praktyce przemysłowej zdarza się, że większe znaczenie od samego wyboru między LSTM a GRU ma poprawne przygotowanie okien czasowych, filtrowanie sygnału i dobór etykiet odpowiadających realnym stanom procesu.

W ostatnich latach część zadań sekwencyjnych przejęły architektury transformatorowe, ale modele rekurencyjne wciąż pozostają cenne, zwłaszcza wtedy, gdy zasoby obliczeniowe są ograniczone, sekwencje mają umiarkowaną długość, a interpretacja dynamiki czasowej jest ważniejsza niż maksymalna skala modelu.

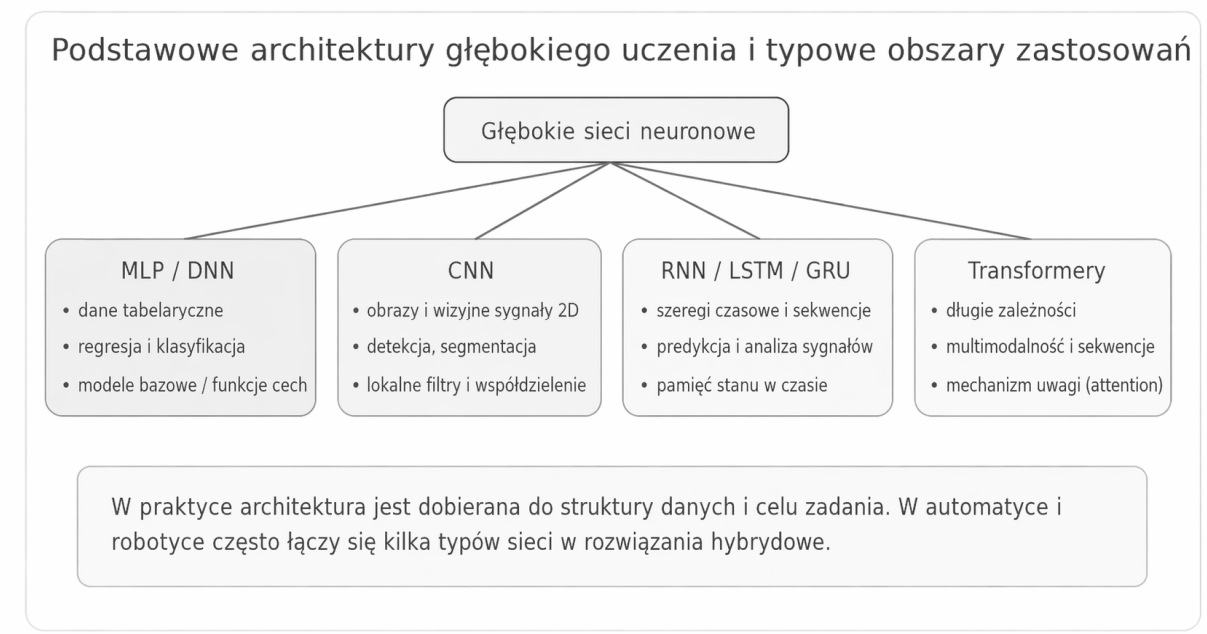
3.11. Mechanizm uwagi i transformery

Mechanizm uwagi, czyli attention, zmienił sposób myślenia o przetwarzaniu sekwencji. Zamiast przechodzić przez dane wyłącznie krok po kroku i polegać na pojedynczym stanie ukrytym, model może nauczyć się, które elementy wejścia są najważniejsze dla interpretacji danego fragmentu. Transformery rozwijają tę ideę, wykorzystując mechanizmy samouwagi do modelowania relacji pomiędzy elementami sekwencji. Dzięki temu mogą skutecznie uchwycić zależności długiego zasięgu i przetwarzać dane w sposób bardziej równoległy niż klasyczne sieci rekurencyjne.

Choć transformery zyskały największy rozgłos w przetwarzaniu języka naturalnego, ich zastosowania szybko rozszerzyły się na wizję komputerową, analizę multimodalną, przetwarzanie sygnałów i systemy robotyczne. W robotyce szczególnie interesujące są dlatego, że umożliwiają integrację informacji pochodzących z wielu źródeł: obrazu, języka, trajektorii, sił kontaktu i danych proprioceptywnych. W automatyce ich potencjał wiąże się z analizą

długich sekwencji procesowych, wykrywaniem zależności rozproszonych w czasie oraz budową modeli predykcyjnych, które potrafią uwzględniać kontekst o dużym horyzoncie czasowym.

Transformery nie są jednak rozwiązaniem darmowym. Wymagają znacznych zasobów obliczeniowych, dużej liczby danych lub dobrego wstępnego uczenia, a ich wdrażanie w systemach czasu rzeczywistego może być trudne. Dla zastosowań inżynierskich oznacza to konieczność świadomego kompromisu między jakością predykcji, kosztem obliczeniowym i niezawodnością. Dlatego należy rozumieć transformery jako bardzo silne narzędzie do wybranych klas problemów, a nie jako automatyczny zamiennik wszystkich wcześniejszych architektur.



Rys. 12. Najczęściej stosowane klasy architektur głębokich i odpowiadające im typy zadań

Źródło: Opracowanie własne.

3.12. Transfer learning, pretrening i uczenie samonadzorowane

Jednym z najważniejszych powodów praktycznego sukcesu współczesnego głębokiego uczenia jest transfer learning, czyli wykorzystanie wiedzy nabytej przez model w jednym zadaniu do rozwiązania zadania pokrewnego. W najprostszym wariancie oznacza to użycie wcześniej wytrenowanej sieci jako ekstraktora cech lub punktu startowego do dalszego dostrajania.

W zadaniach wizyjnych i multimodalnych ma to ogromne znaczenie, ponieważ zebranie dużego, dobrze opisanego zbioru przemysłowego jest często kosztowne albo wręcz niemożliwe. Zastosowanie modelu wstępnie wyuczonego na dużych danych pozwala skrócić czas treningu, zmniejszyć wymagania dotyczące liczby etykiet oraz poprawić jakość rozwiązania.

Coraz większą rolę odgrywa również uczenie samonadzorowane. W tym podejściu model nie otrzymuje klasycznych etykiet od człowieka dla każdej próbki, lecz buduje zadania pomocnicze wynikające z samej struktury danych. Może na przykład uczyć się rekonstrukcji fragmentów wejścia, przewidywania części sekwencji albo zgodności pomiędzy różnymi widokami tego samego obiektu. Tak wyuczone reprezentacje są później wykorzystywane w zadaniach właściwych, takich jak klasyfikacja, segmentacja czy sterowanie.

Z inżynierskiego punktu widzenia jest to bardzo ważny kierunek, ponieważ w praktyce danych nieopisanych jest zwykle znacznie więcej niż danych z etykietami. W automatyce zakłady produkcyjne generują olbrzymie ilości przebiegów pomiarowych, obrazów i logów, ale tylko niewielka część z nich jest dokładnie opisana przez ekspertów. Metody pretreningu i samonadzorowania pozwalają lepiej wykorzystać ten zasób i stanowią pomost między klasycznym uczeniem nadzorowanym a realnymi warunkami przemysłowymi.

3.13. Ocena modeli głębokich i analiza błędów

Ocena jakości modelu głębokiego nie może sprowadzać się do jednej liczby podanej na końcu eksperymentu. W zależności od zadania znaczenie mają inne miary: dokładność klasyfikacji, czułość, precyzja, miara F1, pole pod krzywą ROC, błąd średniokwadratowy, błąd bezwzględny, kalibracja niepewności czy czas odpowiedzi modelu. W automatyce i robotyce bardzo często ważniejsze od maksymalnego wyniku średniego są zachowanie na przypadkach granicznych, odporność na zakłócenia oraz zgodność z wymaganiami czasowymi systemu.

Szczególnie wartościowa jest analiza błędów. Należy sprawdzić, które klasy są najczęściej mylone, w jakich warunkach model traci skuteczność czy błędy koncentrują się przy określonych poziomach zakłóceń, prędkości ruchu, typach materiału albo stanach maszyny. Tylko taka analiza pozwala zdecydować, czy problemem jest architektura, zbiór danych, procedura etykietowania, niewłaściwa miara oceny czy może samo sformułowanie zadania. W praktyce przemysłowej ten etap jest często ważniejszy niż różnice rzędu pojedynczych punktów procentowych pomiędzy modelami.

Nie można również zapominać o walidacji zewnętrznej. Model oceniany wyłącznie na danych pochodzących z jednego stanowiska, jednej kamery albo jednego egzemplarza maszyny może wydawać się znakomity, a po wdrożeniu okazać się niestabilny. Dobra praktyka polega na testowaniu na danych z różnych zmian produkcyjnych, różnych warunków oświetlenia, różnych operatorów i różnych stanów obciążenia procesu.

3.14. Wyjaśnialność, bezpieczeństwo i zaufanie do modelu

Jednym z głównych zarzutów wobec głębokich sieci neuronowych jest ich ograniczona interpretowalność. W wielu przypadkach użytkownik otrzymuje bardzo dobrą predykcję, ale nie ma prostego wglądu w pełny tok rozumowania modelu. W zadaniach czysto konsumenckich bywa to akceptowalne, natomiast w automatyce, medycynie, transporcie czy robotyce przemysłowej może stanowić istotne ograniczenie. Operator, inżynier procesu albo audytor bezpieczeństwa musi wiedzieć, czy model podejmuje decyzje na podstawie sensownych przesłanek, czy opiera się na artefaktach danych.

Z tego powodu rozwija się cały obszar metod wyjaśnialnej sztucznej inteligencji. Obejmuje on mapy aktywacji, analizy wrażliwości, interpretacje lokalne, metody oparte na perturbacjach
i różne sposoby śledzenia wpływu cech wejściowych na decyzję modelu. Narzędzia te nie rozwiązują wszystkich problemów interpretowalności, ale pomagają ocenić, czy model koncentruje się na właściwych fragmentach obrazu, sygnału lub sekwencji. W praktyce przemysłowej stanowią cenne wsparcie walidacji i rozmowy z ekspertami dziedzinowymi.

Z wyjaśnialnością silnie wiąże się bezpieczeństwo. W systemach robotycznych model nie może być oceniany wyłącznie przez pryzmat skuteczności średniej. Należy brać pod uwagę również konsekwencje błędu, możliwość wykrycia niepewności predykcji, mechanizmy przejścia do trybu bezpiecznego oraz relację pomiędzy modelem uczącym się a nadzorującą logiką systemu. W zastosowaniach krytycznych dobrym rozwiązaniem bywają architektury hybrydowe, w których model głęboki dostarcza predykcji lub oszacowania stanu, ale ostateczne decyzje podlegają dodatkowym ograniczeniom i regułom bezpieczeństwa.

3.15. Wdrożenie modeli na brzegu systemu i w czasie rzeczywistym

Różnica między udanym eksperymentem badawczym a udanym rozwiązaniem przemysłowym najczęściej ujawnia się na etapie wdrożenia. Model, który osiąga bardzo dobre wyniki na stacji roboczej z mocnym procesorem graficznym, może nie spełnić wymagań po

przeniesieniu na komputer brzegowy, sterownik przemysłowy lub platformę mobilnego robota. W praktyce trzeba uwzględnić opóźnienie obliczeń, zużycie pamięci, pobór energii, stabilność działania i łatwość integracji z istniejącą architekturą systemu.

Z tego powodu duże znaczenie mają techniki kompresji modelu, kwantyzacji, przycinania wag, destylacji wiedzy oraz dobór architektury pod kątem ograniczeń sprzętowych. Czasami mniejszy model o nieco niższej dokładności, ale gwarantujący powtarzalny czas odpowiedzi, okazuje się znacznie lepszym wyborem niż model cięższy i bardziej dokładny w laboratorium.

W robotyce mobilnej i systemach czasu rzeczywistego takie kompromisy są regułą, a nie wyjątkiem. Nie można też pomijać monitorowania po wdrożeniu. Model powinien być obserwowany pod kątem dryfu danych, zmian rozkładu wejść, pogarszania jakości predykcji i awarii integracyjnych. W praktyce oznacza to potrzebę logowania wyników, okresowego testowania, procedur aktualizacji i zachowania wersji modelu. W inżynierii jest to naturalny element cyklu życia systemu technicznego i sieci neuronowe nie stanowią tutaj wyjątku.

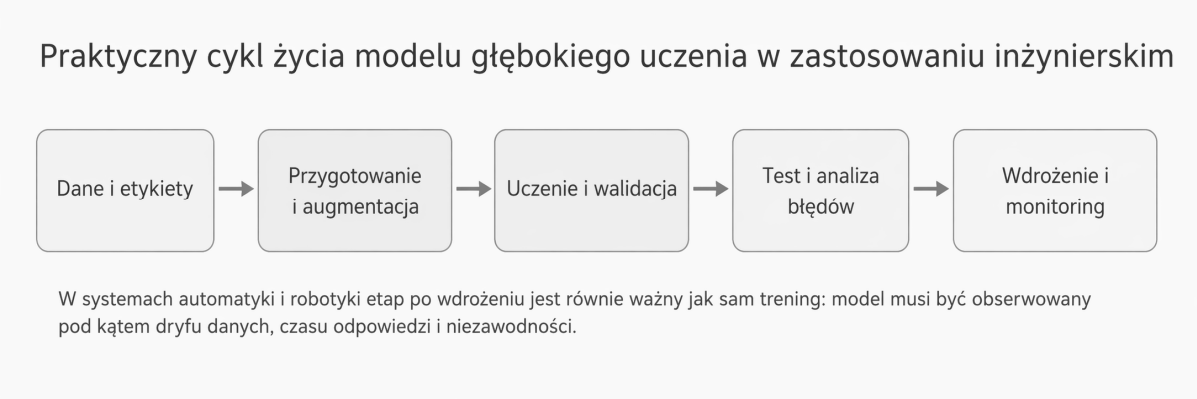
3.16. Projektowanie systemu głębokiego uczenia krok po kroku

Budowa poprawnego systemu głębokiego uczenia wymaga uporządkowanego toku postępowania. Punktem wyjścia musi być definicja problemu inżynierskiego: trzeba ustalić, co dokładnie ma być przewidywane, jakie są dopuszczalne błędy, jaki jest horyzont czasowy predykcji i czy model będzie działał offline, czy w czasie zbliżonym do rzeczywistego. Następnie należy ustalić, jakie dane są dostępne, czy są wiarygodne, jak zostały zebrane oraz czy odpowiadają rzeczywistym warunkom pracy. Dopiero po tej analizie sensowne staje się projektowanie architektury, funkcji straty i procedury walidacji.

Kolejnym etapem jest przygotowanie eksperymentu uczącego. Obejmuje ono podział danych, określenie procedur augmentacji, wybór miar oceny, ustalenie strategii strojenia hiperparametrów oraz zaplanowanie sposobu monitorowania treningu. W praktyce niezwykle pomocne jest prowadzenie eksperymentów porównawczych z modelami bazowymi. Dobrą praktyką jest rozpoczęcie od architektury prostszej i dopiero później zwiększanie złożoności, jeśli dane wskazują, że jest to uzasadnione. Taki tok pracy zmniejsza ryzyko budowania modelu zbyt ciężkiego względem dostępnych danych i zasobów obliczeniowych.

Po wytrenowaniu modelu nie wystarczy podać jednej wartości dokładności. Należy przeanalizować błędy, sprawdzić zachowanie modelu na trudnych przypadkach, ocenić

odporność na zakłócenia i zmienność danych wejściowych oraz upewnić się, że wyniki są zgodne z realnymi wymaganiami aplikacji. W systemach robotycznych i przemysłowych szczególne znaczenie mają ograniczenia czasowe, niezawodność oraz bezpieczeństwo. Model, który działa doskonale w środowisku laboratoryjnym, może okazać się niewystarczający po wdrożeniu na rzeczywiste urządzenie, jeśli nie uwzględniono opóźnień, zmian oświetlenia, dryfu czujników, nietypowych stanów pracy lub ograniczeń mocy obliczeniowej.



Rys. 13. Praktyczny cykl życia modelu głębokiego uczenia w zastosowaniu inżynierskim

Źródło: Opracowanie własne.

3.17. Zastosowania w automatyce i robotyce

Zakres zastosowań głębokiego uczenia w automatyce i robotyce szybko się rozszerza, ale warto widzieć go przez pryzmat konkretnych funkcji inżynierskich. Pierwszą z nich jest percepcja. Sieci konwolucyjne, segmentacyjne i wielomodalne umożliwiają rozpoznawanie obiektów, śledzenie elementów ruchomych, wykrywanie wad powierzchni, analizę sceny robota mobilnego, identyfikację części na linii montażowej oraz estymację położenia i orientacji. Tam, gdzie dawniej projektowano rozbudowane zestawy ręcznych cech obrazowych, współczesne systemy coraz częściej uczą się reprezentacji bezpośrednio z danych.

Drugim dużym obszarem jest diagnostyka i utrzymanie predykcyjne. Głębokie sieci uczą się rozpoznawać wzorce degradacji urządzeń na podstawie sygnałów drganiowych, akustycznych, termicznych lub elektrycznych. Dzięki temu możliwe jest wcześniejsze wykrywanie zużycia łożysk, uszkodzeń przekładni, nieszczelności, niewyważenia czy niestabilnych stanów pracy napędów. W systemach przemysłowych takie rozwiązania mogą zmniejszać liczbę awarii nieplanowanych i lepiej wspierać decyzje serwisowe. Trzeba jednak

pamiętać, że najtrudniejsze są zwykle nie same algorytmy, lecz pozyskanie wiarygodnych danych z rzadkich stanów uszkodzeń oraz poprawna interpretacja wyników modelu.

Trzecim ważnym zastosowaniem jest modelowanie i sterowanie. Sieci neuronowe mogą pełnić funkcję modeli zastępczych obiektu, estymatorów zakłóceń, predyktorów przebiegu procesu albo komponentów sterowania opartego na uczeniu. W robotyce pojawiają się jako elementy planowania chwytania, przewidywania trajektorii, sterowania manipulatorem, lokomocji czy nawigacji. Szczególnie obiecujące są układy hybrydowe, w których głęboki model współpracuje z klasycznym regulatorem albo z metodą optymalizacji ruchu. Takie połączenie pozwala wykorzystywać siłę danych, ale jednocześnie zachować inżynierską kontrolę nad bezpieczeństwem i ograniczeniami układu.

3.18. Ograniczenia i odpowiedzialne stosowanie

Mimo spektakularnych sukcesów uczenie głębokie nie jest metodą pozbawioną ograniczeń. Po pierwsze, wymaga zwykle dużej ilości danych lub przynajmniej danych reprezentatywnych względem warunków rzeczywistych. Po drugie, bywa kosztowne obliczeniowo, co ma znaczenie przy wdrożeniach na sterownikach, komputerach brzegowych i systemach czasu rzeczywistego. Po trzecie, interpretowalność wielu modeli pozostaje ograniczona. Z punktu widzenia inżynierii oznacza to, że nie zawsze można bezpośrednio wyjaśnić, dlaczego model podjął daną decyzję. Problem ten jest szczególnie istotny w diagnostyce, bezpieczeństwie funkcjonalnym i systemach autonomicznych.

Kolejnym ryzykiem jest wrażliwość na przesunięcie rozkładu danych, czyli sytuację, w której dane pojawiające się po wdrożeniu różnią się od danych treningowych. W praktyce przemysłowej może to wynikać ze zmiany czujnika, ustawień maszyny, rodzaju materiału, warunków środowiskowych albo starzenia się aparatury. Model, który został wytrenowany w jednym kontekście, nie musi zachować tej samej jakości w innym. Dlatego tak ważne jest monitorowanie jakości po wdrożeniu, okresowa rewalidacja oraz świadome zarządzanie cyklem życia modelu.

W odpowiedzialnym stosowaniu głębokiego uczenia liczy się zatem nie tylko skuteczność laboratoryjna, lecz również przejrzystość procedury, powtarzalność eksperymentów, dokumentacja danych, analiza błędów i gotowość do aktualizacji modelu. W zastosowaniach krytycznych nie należy traktować sieci neuronowej jako autonomicznego bytu, któremu wystarczy „zaufać”. Lepiej myśleć o niej jako o komponencie systemu

technicznego, który musi być oceniany z punktu widzenia niezawodności, bezpieczeństwa i zgodności z wymaganiami aplikacji.

3.19. Podsumowanie

Sztuczne sieci neuronowe i uczenie głębokie tworzą dziś jeden z najważniejszych filarów współczesnej sztucznej inteligencji. Ich siła wynika z umiejętności budowy złożonych reprezentacji danych i modelowania relacji, których opis analityczny bywa trudny albo niepraktyczny. Dla studenta automatyki i robotyki kluczowe jest jednak nie tyle zapamiętanie nazw poszczególnych architektur, ile zrozumienie ich wspólnej logiki: wejścia są przekształcane warstwa po warstwie, model uczy się przez minimalizację funkcji straty, a końcowa jakość zależy jednocześnie od danych, architektury, optymalizacji i walidacji.

Najważniejszy wniosek praktyczny jest taki, że uczenie głębokie nie zastępuje myślenia inżynierskiego. Przeciwnie, wymaga go w jeszcze większym stopniu. Żeby zbudować dobry model, trzeba rozumieć obiekt, proces zbierania danych, ograniczenia sprzętowe, charakter błędów i wymagania wdrożeniowe. Sieci neuronowe są bardzo silnym narzędziem, ale dopiero wtedy, gdy zostaną włączone do poprawnie zaprojektowanego procesu inżynierskiego. W tym sensie uczenie głębokie nie jest alternatywą dla rzetelnej analizy technicznej, lecz jej nowoczesnym rozwinięciem.

3.20. Zestawienie najważniejszych architektur

Dla uporządkowania materiału warto zestawić najważniejsze typy architektur z punktu widzenia danych wejściowych, mocnych stron i typowych zastosowań. Tabela nie zastępuje pełnego opisu, ale ułatwia szybkie porównanie rozwiązań spotykanych najczęściej w praktyce inżynierskiej i dydaktyce.

Tabela 6. Najważniejsze typy architektur z punktu widzenia danych wejściowych i zastosowań

Architektura	Typ danych	Najważniejsza zaleta	Przykładowe zastosowanie
MLP / DNN	wektory cech, dane tabelaryczne	prosta implementacja, dobra baza odniesienia	regresja parametrów procesu, klasyfikacja stanów pracy
CNN	obrazy, mapy cech, sygnały 1D/2D	lokalna analiza wzorców i współdzielenie wag	inspekcja wizyjna, segmentacja, analiza sygnałów drganiowych

Architektura	Typ danych	Najważniejsza zaleta	Przykładowe zastosowanie
RNN / LSTM / GRU	szeregi czasowe, sekwencje	modelowanie zależności czasowych i pamięci stanu	predykcja sekwencji, analiza trajektorii, rozpoznawanie zdarzeń
Transformery	długie sekwencje, dane multimodalne	mechanizm uwagi i zależności dalekiego zasięgu	wizja komputerowa, fuzja sensoryczna, modele językowo-robotyczne
Autoenkodery / modele generatywne	dane bez etykiet lub częściowo oznaczone	uczenie reprezentacji, rekonstrukcja, detekcja anomalii	odszumianie, kompresja, monitorowanie stanu procesu

Bibliografia

Holland, J. H., Adaptation in Natural and Artificial Systems, MIT Press, Cambridge, MA, 1992 (wyd. oryg. 1975).

Goldberg, D. E., Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, MA, 1989.

Michalewicz, Z., Algorytmy genetyczne + struktury danych = programy ewolucyjne, WNT, Warszawa, 1999.

Eiben, A. E., Smith, J. E., Introduction to Evolutionary Computing, 2nd ed., Springer, Berlin–Heidelberg, 2015.

Bäck, T., Fogel, D. B., Michalewicz, Z. (red.), Handbook of Evolutionary Computation, IOP Publishing / Oxford University Press, 1997.

Beyer, H.-G., Schwefel, H.-P., "Evolution Strategies: A Comprehensive Introduction," Natural Computing, 1(1), 2002, s. 3–52.

Storn, R., Price, K. V., "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," Journal of Global Optimization, 11, 1997, s. 341–359.

Koza, J. R., Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT Press, Cambridge, MA, 1992.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," IEEE Transactions on Evolutionary Computation, 6(2), 2002, s. 182–197.

Mitchell, T. M., Machine Learning, McGraw-Hill, New York, 1997.

Bishop, C. M., Pattern Recognition and Machine Learning, Springer, New York, 2006.

Hastie, T., Tibshirani, R., Friedman, J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer, New York, 2009.

Alpaydin, E., Introduction to Machine Learning, 4th ed., MIT Press, Cambridge, MA, 2020.

James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J., An Introduction to Statistical Learning: with Applications in Python, Springer, Cham, 2023.

Murphy, K. P., Probabilistic Machine Learning: An Introduction, MIT Press, Cambridge, MA, 2022.

Russell, S., Norvig, P., Artificial Intelligence: A Modern Approach, 4th ed., Pearson, Harlow, 2021.

Goodfellow, I., Bengio, Y., Courville, A., Deep Learning, MIT Press, Cambridge, MA, 2016.

LeCun, Y., Bengio, Y., Hinton, G., "Deep Learning," Nature, 521(7553), 2015, s. 436–444.

Schmidhuber, J., "Deep Learning in Neural Networks: An Overview," Neural Networks, 61, 2015, s. 85–117.

Szeliski, R., Computer Vision: Algorithms and Applications, 2nd ed., Springer, 2022.

Zhang, A., Lipton, Z. C., Li, M., Smola, A. J., Dive into Deep Learning, Cambridge University Press, 2023.

Projektowanie linii zautomatyzowanych

(prof. dr hab. inż. Andrzej Milecki)

Wprowadzenie

Automatyzacja masowych procesów produkcyjnych została po raz pierwszy zastosowana na początku lat 50. XX wieku w firmie Ford. Wprowadzono w niej ruchomą linię montażową, która znacznie zwiększyła szybkość i wydajność produkcji. Ta rewolucyjna innowacja polegała także na tym, że każdy pracownik wykonywał wielokrotnie to samo zadanie. Takie podejście usprawniło produkcję i umożliwiło zwiększenie wydajności. W latach 1920–1921 obniżono cenę samochodu Forda do 355 dolarów, a wielkość produkcji osiągnęła 1 250 000 sztuk.

W latach 60. XX w. elektronika osiągnęła na tyle wysoki poziom rozwoju, że przestała być przedmiotem zainteresowania wyłącznie twórców radioodbiorników, telewizorów i komputerów budowanych na bazie układów cyfrowych TTL średniej skali integracji, a stała się bardzo interesująca dla konstruktorów i wytwórców urządzeń przemysłowych. To wtedy rozwinęły się np. serwonapędy prądu stałego a za nimi obrabiarki sterowane numerycznie, roboty oraz zautomatyzowane i skomputeryzowane linie produkcyjne. Układy i elementy pomiarowe bezpośrednio współpracowały z urządzeniami elektronicznymi. Elektronika zaczęła pojawiać się w samochodach i maszynach roboczych, a coraz większa rzesza hobbystów zaczęła budować różnorodne drobne urządzenia elektroniczne, typu zamki, sterowniki oświetlenia, zabawki itp. To wtedy także pojawiły się poważne konstrukcje, w których zastosowano komputery do sterowania maszynami.

Najważniejsze postępy w automatyzacji zostały dokonane dzięki rozwojowi, w zakresie elektroniki i automatyzacji. Ich wprowadzenie do budowanych urządzeń stało się najczęściej stosowaną metodą prowadzącą do poprawy ich właściwości i funkcjonalności. W drugiej połowie lat 70. XX w. zaczęto produkować mikroprocesory, a na początku lat 80. Mikrokontrolery, które były miniaturowymi sterownikami. Dzięki miniaturyzacji możliwe stało się ich stosowanie w niemalże dowolnym urządzeniu. Oznaczało początek używania bardziej zaawansowanych technik automatyzacji w przemyśle.

W wytwarzanych już od połowy lat 70. XX w. urządzeniach powszechna zaczęła być koegzystencja mechaniki z elektroniką. Wtedy to powstała nowa technika projektowania,

nazywana mechatroniką. Równolegle rozwijana była automatyzacja całych linii produkcyjnych, na których coraz częściej zaczęły pojawiać się roboty przemysłowe. Automatyzacja linii produkcyjnych i to warunek rozwoju oraz utrzymania się na rynku każdej z firm. Zautomatyzowane linie produkcyjne cechują się wysoką powtarzalnością i dokładnością. Bez tego trudno jest uzyskać najwyższą jakość oferowanych produktów. Czynności powtarzalne, monotonne można zautomatyzować za pomocą robotów, które wykonują je o wiele szybciej i dokładniej niż człowiek. Skróceniu ulega czas, niezbędny do wykonania określonych działań i wytworzenia produktu.

Rozwój techniki mikroprocesorowej doprowadził do nowych możliwości przetwarzania i przepływu informacji w sterowanych urządzeniach. Zalety mikroprocesora, takie jak miniaturowe rozmiary, duże moce obliczeniowe, łatwość programowania oraz niska cena pozwoliły na łatwe ich wbudowywanie do urządzeń. Mikrokontrolery są używane w niemalże każdym urządzeniu, np. w robocie, podajniku, podnośniku oraz w maszynach: tokarkach, wiertarkach, frezarkach, prasach, krawędziarkach itp. Dzisiaj prawie każde zaawansowane technicznie urządzenie, ma przynajmniej prosty moduł elektroniczny z maleńkim mikrokontrolerem wykonującym pomiary oraz sterującym silnikami i innymi elementami wykonawczymi.

Prawdziwą rewolucją w automatyce było upowszechnienie się komputerów klasy PC. Po odpowiednim dostosowaniu do warunków przemysłowych, zaczęły one służyć zarówno do nadzorowania układów sterowania, jak i do sterowania zaawansowanymi stacjami i całymi liniami produkcyjnymi. Wraz z ich wprowadzeniem automatyka uzyskała nowe możliwości, a zakres automatyzacji poszerzył się o zupełnie nowe obszary zastosowań, związane z kompleksową automatyzacją produkcji oraz z elastycznym zarządzaniem i sterowaniem przedsiębiorstwem. Dzisiaj trudno jest spotkać nowoczesną linię produkcją bez sterowania komputerowego.

1. Podstawy automatyzacji urządzeń

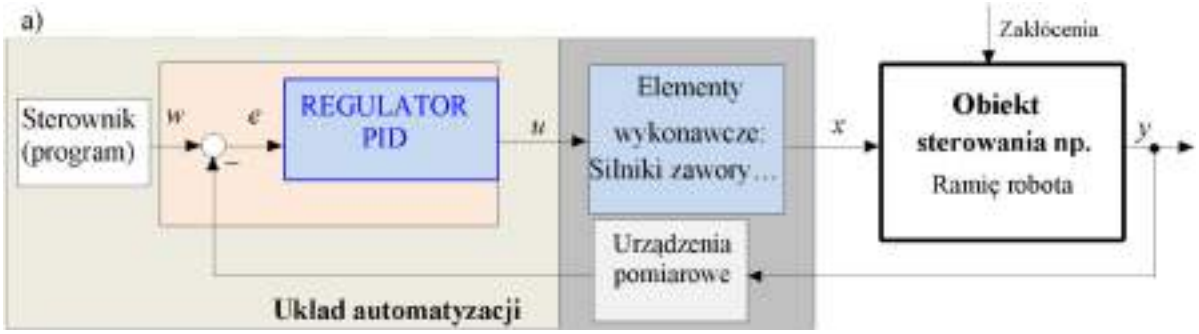
Automatyzacja w praktyce sprowadza się do zastosowania urządzeń pomiarowych i napędów oraz sterowników, dzięki którym urządzenia albo procesy produkcyjne będą działały samoczynnie, bez albo z ograniczonym udziałem ludzi – pracowników. W celu automatyzacji stosowane są następujące elementy i/albo urządzenia:

1) pomiarowe (czujniki, przetworniki oraz zespoły pomiarowe)

2) wykonawcze (silniki, zawory, zasuwy, elektromagnesy, siłowniki, zespoły napędowe, pompy, podajniki, regulatory)
3) części centralnej i sterującej (panele operacyjne, przełączniki, nastawniki, wskaźniki, wyświetlacze)
4) sterowniki (PLC, regulatory, komputery przemysłowe – IC, moduły komunikacji).

Pierwsze z nich przeznaczone są do mierzenia wielkości wyjściowych z poszczególnych, sterowanych podzespołów. Najczęściej mierzone są wielkościami mechaniczne takie jak: siła, przyspieszenie, prędkość, położenie, ciśnienie, temperatura itp. Do tego stosowane są różne czujniki i urządzenia pomiarowe, które generują na wyjściach sygnały elektryczne, czyli napięcie *U* albo natężenie prądu *I* a niekiedy impulsy. Definicja mówi, że sygnał to dowolna wielkość fizyczna będąca funkcją (zmienną) w czasie. Ze względu na rodzaj sygnału, w automatyzacji wyróżnia się sygnały elektryczne:

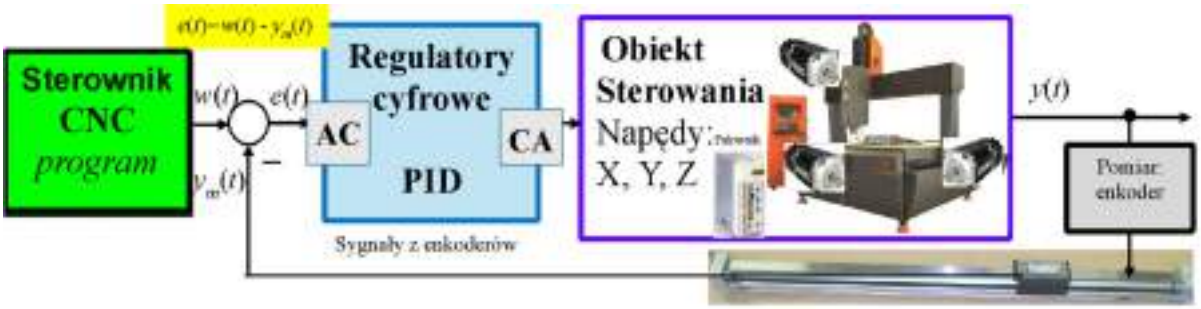
- analogowe (ciągłe) – wartość sygnału określona jest przez wartości wielkości fizycznej i może przyjmować dowolną wartość z określonego przydziału, zwykle: napięciowe –10V do +10V, 0 do 10V, 0 do 5V albo prądowe 0 do 20 mA, 4 do 20 mA)
- binarne (przełączające, dwustanowe) – sygnał może przyjmować tylko jedną z dwóch wartości: 0 lub 1, kodowaną odpowiednio jako: 0V albo 24VDC.



Rys. 1. Zamknięty układ automatyzacji schemat blokowy: *w* – sygnał zadany, *e* – uchyb regulacji, *x* – sygnał sterujący, *y* – sygnał wyjściowy

W zależności od tego, jakie rodzaj sygnałów jest wykorzystywany w procesie sterowania, mówimy o sterowaniu analogowym (ciągłym) albo o sterowaniu binarnym. W pierwszym przypadku powszechnie stosuje się układy sterowania ze sprzężeniem zwrotnym, polegającym na pomiarze rzeczywistej wartości sygnału *y* na wyjściu sterowanego obiektu i przekazaniu go na wejście, w celu porównania z wartością sygnału zadanego *w* (rys. 1). Porównanie to polega na wyznaczeniu różnicy między tymi sygnałami. Uzyskany wynik $e = w - y$ nazywa się błędem albo uchybem regulacji. Wykorzystuje się go do skorygowania oddziaływania regulatora

sterującego obiektem, w taki sposób, aby zminimalizować ten uchyb, a w ostateczności doprowadzić go do zera. Taki układ nazywa się zamkniętym układem regulacji z ujemnym sprzężeniem zwrotnym (rys. 1). Sygnały wyjściowe są mierzone za pomocą różnego rodzaju czujników i urządzeń pomiarowych. Od ponad 100. lat, do regulacji stosowane są głównie regulatory proporcjonalno-całkująco-różniczkujące (ang. *Proportional-Integral-Derivative* – PID). Wielkości fizyczne, które stanowią wyjście automatyzowanego urządzenia są przedmiotem oddziaływania układu regulacyjnego. Nazywamy je wielkościami bądź sygnałami regulowanymi *y*, np.: siła, ciśnienie, przyspieszenie, prędkość, przemieszczenie, temperatura. Oddziaływanie regulatora na przebieg procesu odbywa się za pośrednictwem elementów wykonawczych. Wartości zadane są najczęściej ustalane przez systemu sterowania nadrzędnego, w którym zapisany jest program działania urządzenia a nawet całej linii produkcyjnej. W układach regulacji ręcznej, rolę regulatora pełni człowiek.



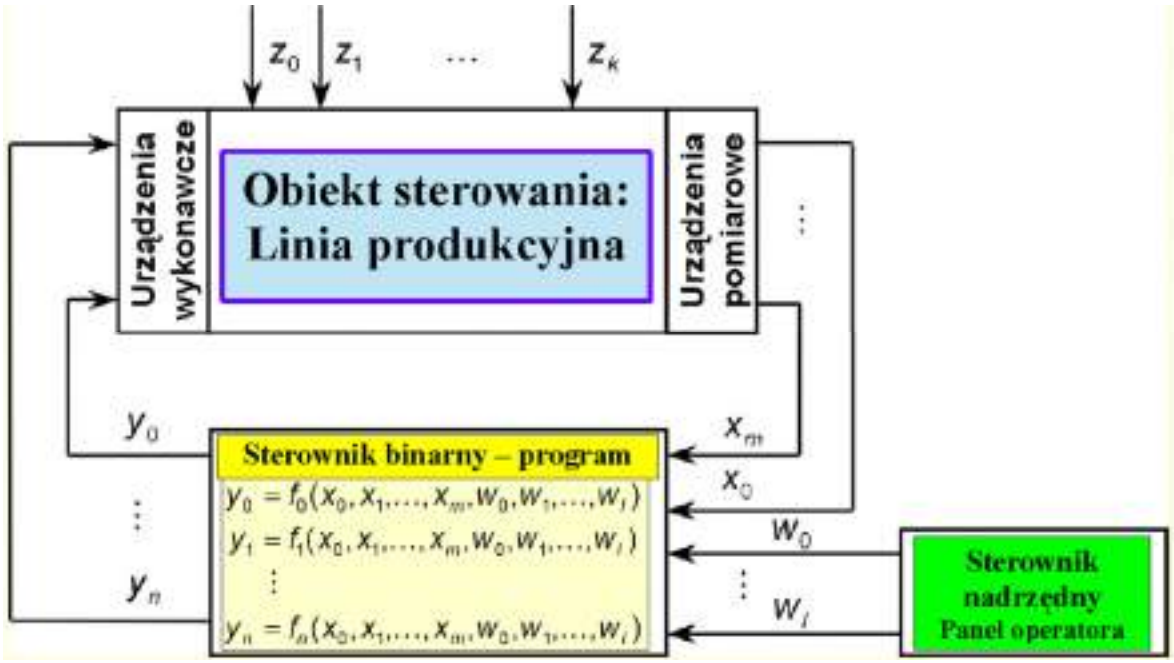
Rys. 2. Zamknięty układ automatyzacji obrabiarki CNC

Na rysunku 2 przedstawiono schemat układu sterowania obrabiarki CNC. W samym sterowniku CNC jest zapisany program obróbki tj. kolejne pozycje na trajektorii i parametry ruchu narzędzia itp. Na tej podstawie sterownik CNC, generuje sygnały zadane do poszczególnych sterowników elektronicznych silników obrabiarki. Przesuwają one elementy obrabiarki, tzn. najczęściej uchwyt z narzędziem obróbkowym np. wiertło, nóż tokarski albo frez. Przemieszczenia w poszczególnych osiach są mierzone za pomocą elementów pomiarowych, których sygnały wyjściowe są podawane jako sprzężenie zwrotne na wejście układu – do węzła odejmującego. Na wejściach i wyjściach regulatora zastosowane są przetworniki Analogowo-Cyfrowe (AC) oraz Cyfrowo-Analogowe (CA), które zamieniają odpowiednio sygnały ciągłe w postaci napięcia na liczby oraz liczby na napięcie.

W praktyce przemysłowej, większość procesów na liniach produkcyjnych, dotyczy jednak wykonywania operacji typu włącz/wyłącz. W związku z tym, w układzie sterowania występują sygnały binarne „0” albo „1”, kodowane napięciem 24VDC (poziom wysoki tzn.

element jest obecny albo włącz) oraz 0V (poziom niski – brak elementu albo wyłącz). Automatyzacji podlega wielowejściowy obiekt, składający się z szeregu urządzeń tj. napędów, transporterów, podajników, zespołów pomiarowych, robotów i chwytaków, obrabiarek, automatów montażowych, pakujących itp. W każdym z tych elementów zastosowane są różne czujniki i układy pomiarowe generujące na wyjściu sygnały binarne 0/1 (jest/nie ma) oraz elementy wykonawcze typu silniki, elektromagnesy, elektrozawory itp.

Do sterowania układów binarnych stosowane są sterowniki przemysłowe, realizujące funkcje boolowskie, określone macierzą $\mathbf{F} = [f_0(x_1, x_2, \ldots x_m, w_1, \ldots), f_1(\ldots), \ldots f_n(\ldots)]$. Zaprogramowany sterownik binarny (ang. *Programmabli Logic Controller* – PLC) odpowiedzialny jest za generowanie zadań sterowania, w postaci sygnałów binarnych. W każdej chwili czasowej (kroku), do sterownika podawany jest wektor wejść $\mathbf{X}=[x_0, x_1, \ldots x_m]$ oraz wektor sygnałów zadanych $\mathbf{W} = [w_1, w_2, \ldots w_l]$, a sterownik wysyła do urządzeń wykonawczych wektor, określający stany wyjść $\mathbf{Y}=[y_1, y_2, \ldots y_n]$. Schemat układu sterowania binarnego pokazano na rys. 3.



Rys. 3. Schemat blokowy automatyzacji binarnej linii produkcyjnej (wszystkie sygnały są binarne, tzn. 0/1): W – wektor sygnałów zadanych, X – wektor sygnałów mierzonych, **F** – wektor funkcji sterowania, **Y** – wektor sygnałów wyjściowych, Z – wektor zakłóceń

Obecnie najczęściej stosowanymi formami automatyzacji są:

- sztywna (Fixed Automation) – stosowana do produkcji dużej ilości danego lub bardzo podobnych do siebie produktów. Zakłada się, że ich produkcja będzie trwała przynajmniej kilka lat i nie będą wprowadzane do jej automatyzacji żadne zmiany. Obecnie, ta forma automatyzacji w zasadzie nie jest stosowana.
- programowalna (Programmable Automation) – zaprojektowana do miejsc, gdzie produkty wytwarza się w dużych seriach. Działanie linii produkcyjnej może być dostosowywane poprzez zmianę programów sterowania, do aktualnych potrzeb firmy.
- procesowa (Process Automation) – stosowana w przemysłach, gdzie produkcja odbywa się w sposób ciągły m.in. w rafineriach, elektrowniach czy zakładach przetwórstwa chemicznego. Koncentruje się na monitorowaniu i kontrolowaniu zmiennych procesowych, takich jak temperatura, ciśnienie, przepływ, skład chemiczny itp.
- elastyczna (Flexible Automation) – zaawansowana forma automatyzacji umożliwiająca szybkie dopasowanie produkcji do aktualnie wytwarzanych elementów bez żadnych przestojów. Przykładem tego typu rozwiązania są np. systemy produkcji just-in-time w produkcji samochodów.
- zintegrowana (Integrated Automation) – polega na integracji wszystkich systemów przedsiębiorstwa od zarzadzania, przez projektowanie, produkcję, aż po dystrybucję. Jej przykładem są systemy zarządzania wszystkimi zasobami przedsiębiorstwa typu ERP (Enterprice Resource Planning) albo systemy sterowania cyklem życia produktu (PLM).

W automatyzacji sztywnej algorytm sterowania zostaje zapisany w sposób trwały w postaci połączeń elektrycznych poszczególnych elementów elektrycznych i w przypadku konieczności zmiany programu, należy zmienić te połączenia oraz zastosować nowe elementy w układ sterowania. Przykładem może być układ przekaźnikowy albo scalony typu, płytka z elementami elektronicznymi i/albo elementy elektromechaniczne. Zmienno-programowe układy sterowania tzn. automatyzacja programowalna, procesowa i zintegrowana, to układy sterujące, których algorytm sterowania został zapisany w sposób nietrwały w pamięci urządzenia sterującego i w przypadku konieczności zmiany programu należy tylko zmienić zawartość pamięci układu.

System automatyki to najbardziej rozbudowane rozwiązanie techniczne w zakresie automatyzacji. Składa się z wielu elementów, układów i urządzeń automatyki połączonych ze sobą w jedną całość – system. Realizuje on kompleksowo problem sterowania całym zakładem albo jego złożonym fragmentem, np. linią do produkcji karoserii albo stacją obróbczą itp.

2. Elementy pomiarowe stosowane w automatyzacji

Wprowadzanie automatyzacji do sterowania linią produkcyjną wymaga zastosowania na niej różnorodnych elementów automatyzacji. Na rynku dostępna jest obecnie cała ich gama i konieczna jest ich znajomość, aby właściwe je dobrać i zaprojektować układ automatyzacji linii produkcyjnej. Dzięki osiągnięciom mikroelektroniki powstały także różnorodne elementy wykonawcze. Elementy automatyki to człony spełniające w układzie automatyki bądź w urządzeniu automatyzowanym proste funkcje, takie jak: zmiana postaci sygnałów, ich wzmocnienie oraz porównanie. Elementem jest więc czujnik wykrywający albo pomiarowy, wzmacniacz sygnału, elektrozawór, silnik, prądnica, grzałka itp., a także dowolny regulator albo sterownik.

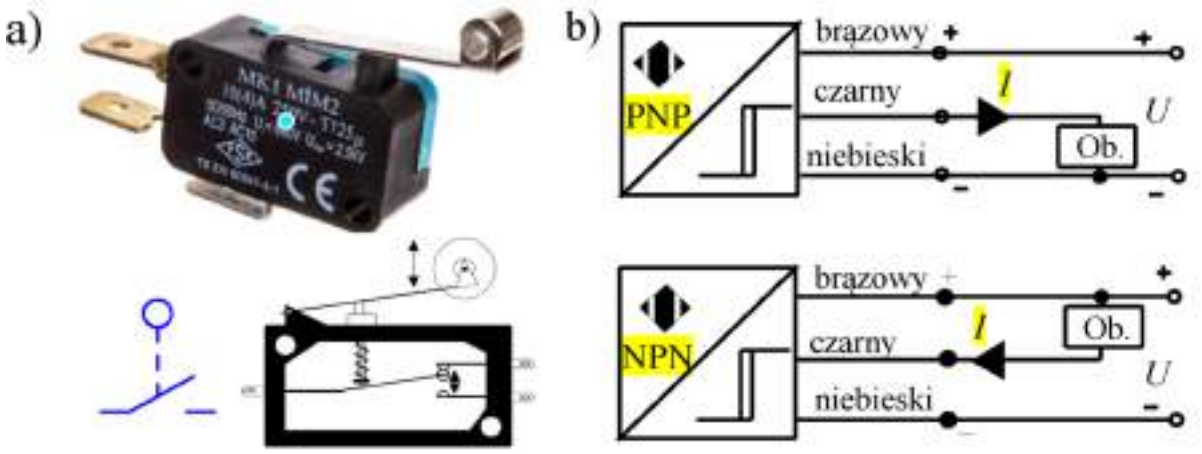
Człony spełniające funkcje bardziej złożone nazywane są urządzeniami automatyki, np. urządzenia pomiarowe (zawierające czujniki, przetworniki, wzmacniacze, filtry itp.), urządzenia napędowe zawierające: sterownik-falownik, czujnik pomiaru położenia, silnik, przekładnie, prowadnice itp. Układ automatyki powstaje z połączenia elementów i urządzeń w pewien zespół, wykonujący określone zadanie. Możemy przykładowo rozróżnić układ sterowania robotem, obrabiarką itp. Z kolei system automatyki to kompletny zestaw elementów, urządzeń i układów, czyli zestaw aparatury kontrolno-pomiarowej, które zastosowano do automatyzacji przynajmniej np. stacji obróbczej. Zestaw taki, aby mógł być nazwany systemem, musi być na tyle duży i kompletny, aby umożliwiał sterowanie liniami produkcyjnymi albo wręcz całymi fabrykami. Systemy do automatyzacji linii przemysłowych wymagają zastosowania czujników do wykrywania obiektów, pomiaru parametrów procesu oraz szeregu różnych napędów elektrycznych, pneumatycznych i hydraulicznych. Do sterowania stosowane są różne sterowniki, połączone w sieć. Odpowiednio dobrane mogą sprostać wielu wyzwaniom stając się niezawodnym rozwiązaniem w wielu aplikacjach.

Czujniki dyskretne on/off (0/1)

Czujniki przemysłowe to elementy układu automatyki, które przeznaczone są do wykrywania i rejestrowania sygnałów z urządzeń na liniach produkcyjnych. W binarnych układach automatyzacji stosowane są różnego rodzaju czujniki wykrywające obecność lub brak, jakiegoś wybranego elementu np. obecność butelki, nakrętki, śruby itp., bądź parametru np. przekroczenie ciśnienia 8 Pa albo temperatury 50°C. Stosowane są czujniki stykowe (włączniki, wyłączniki) w postaci tzw. krańcówek oraz bezkontaktowe czujniki zbliżeniowe. Najbardziej popularne są czujniki: indukcyjnościowe, pojemnościowe, fotoelektryczne albo ultradźwiękowe. Wykorzystują one różne zasady działania i mogą wykrywać elementy, wykonane z różnych materiałów, ale ich wspólną cechą jest zwieranie i rozwieranie połączenia elektrycznego na wyjściu czujnika, za pomocą styków albo tranzystorów.

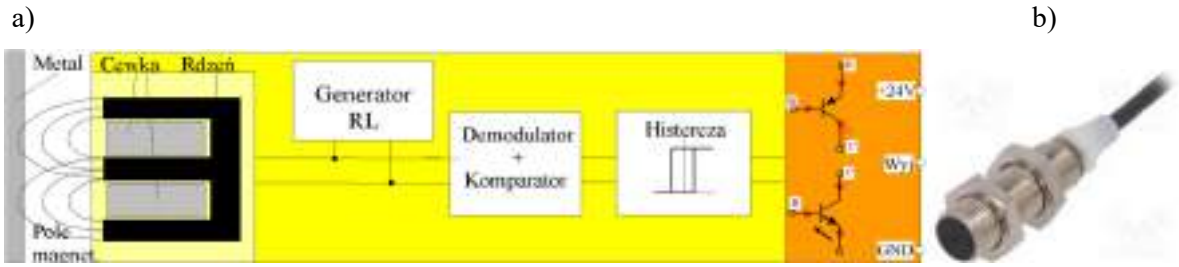
Najstarsze, ale też najtańsze, są czujniki mechaniczne – stykowe, które do zmiany stanu wymagają bezpośredniego przełączenia zestyków przez wykrywany obiekt (rys. 4a). Są stosowane do wykrywania położenia lub obecności obiektów w zadanych położeniach, najczęściej na krańcach ruchu. Stosowane są w nich elementy, które poruszają się w reakcji na kontakt z obiektem. Przykładem może być włącznik albo przycisk aktywowany ręcznie albo przez kontakt mechaniczny np. z wózkiem. Dzięki swojej prostocie są niezawodne, choć ich wadą jest iskrzenie, zużywanie się styków, ich drgania itp.

Dwustanowe wyjście wszystkich bezstykowych, czujników umożliwia bezpośrednią współpracę z przekaźnikami i programowanymi sterownikami logicznymi (PLC). W stopniach wyjściowych tych czujników stosowane są tranzystory typu NPN albo PNP. Każde z tych dwóch typów wyjść jest zwykle zabezpieczone przez nieprawidłowym połączeniem oraz przed przepięciem za pomocą diod. Czujniki są zwykle zasilane napięciem stałym z zakresu od 10 do 30 V. Stosowane w nich tranzystory wyjściowe mogą być obciążone prądem równym maksymalnie 0,2 A. Schemat i sposób połączenia tych czujników pokazano na rys. 4b. Istotny jest kierunek przepływu prądu wyjściowego: w stanie „1” w czujniku PNP prąd wypływa z czujnika a w czujniku PNP wpływa.



Rys. 4. Schematy czujników: a) krańcówki aktywowane przez nacisk (zdjęcie, symbol, budowa), b) bezkontaktowy czujnik indukcyjnościowy z wyjściami tranzystorowymi typu: PNP i PNP

Czujniki indukcyjne są elementami automatyki wykrywającymi wprowadzenie metalu w strefę czułości/działania czujnika. W układach automatyki zastępują one stykowe wyłączniki krańcowe, łączniki drogowe itp. Służą do określenia położenia ruchomych części maszyn i urządzeń. Stosuje się je tam, gdzie są wymagane: duża niezawodność, częstotliwość przełączania, precyzja i powtarzalność działania oraz możliwość w pracy w trudnych warunkach środowiskowych.



Rys. 5. Budowa bezkontaktowego czujnika indukcyjnego: a) schemat, b) widok

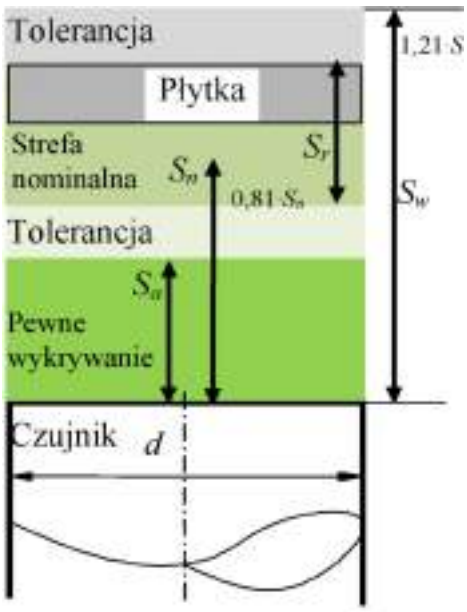
Częścią wykrywającą obiekt jest cewka nawinięta na rdzeniu ferrytowym, połączona z generatorem (rys. 5). Generuje ona w otoczeniu czoła czujnika zmienne pole magnetyczne, które w metalu wytwarza prądy wirowe, co z kolei powoduje „obciążenie” układu oscylatora i w efekcie spadek amplitudy oscylacji. Zmiany te są wykrywane przez komparator i przy pewnej, charakterystycznej dla danego typu czujnika odległości obiektu metalowego od jego czoła, następuje skokowa zmiana stanu (napięcia) na wyjściu komparatora. Strefa działania to podstawowy parametr czujnika bezkontaktowego indukcyjnego (rys. 6). Wynika ona z typu oraz gabarytów czujnika. W katalogach wyrobu i na tabliczkach znamionowych jest podawana

najczęściej strefa nominalna S_n. Wynosi ona zwykle od ok. 0,5 do 1,0 średnicy czujnika. Strefa rzeczywista S_r (uwzględnia fabryczną tolerancję wykonania wyrobu), spełnia warunek:

$$0{,}9\ S_n < S_r < 1{,}1\ S_n. \tag{1}$$

Pomiar strefy rzeczywistej w warunkach fabrycznych polega na zbliżaniu do powierzchni czołowej w osi czujnika kwadratowej płytki ze stali o grubości 1 mm i o boku równym średnicy czujnika. Najistotniejsza jest jednak strefa robocza S_w, która gwarantuje działanie czujnika w pełnym zakresie temperatury i napięcia zasilającego dla długiego okresu eksploatacji. Jest ona kreślona nierównościami:

$$0{,}81\ S_n < S_w < 1{,}21\ S_n. \tag{2}$$



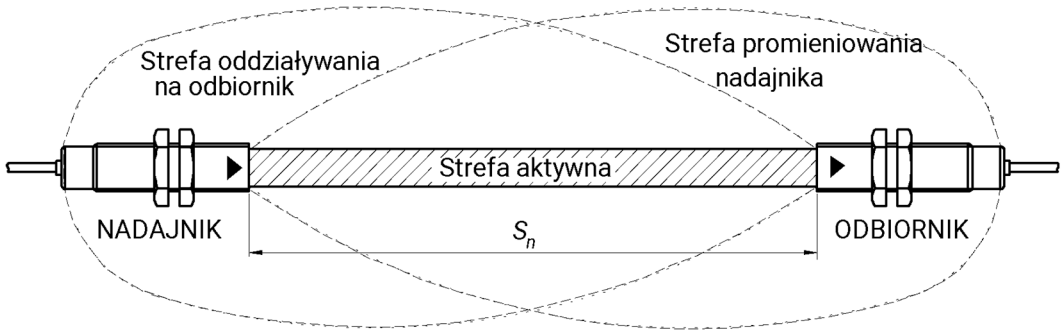
Rys. 6. Strefa działania czujnika

W realnych warunkach eksploatacji strefa działania może różnić się aż o 21% od wartości nominalnej. Strefa działania, tzn. odległość obiektu od czoła czujnika pojemnościowego, przy której następuje przełączenie wyjścia czujnika, zależy także od rodzaju materiału wykrywanego. W katalogach podaje się następujące współczynniki korekcyjne strefy działania: metale – 1,0; szkło – 0,5; PCW – 0,6; woda – 1,0; olej – 0,1. Maksymalna częstotliwość przełączania tych czujników wynosi od ok. 0,5 do 2,5 kHz.

W przemyśle najczęściej stosowane są czujniki indukcyjnościowe, ponieważ większość wykrywanych elementów wykonanych jest ze stali. Do wykrywania materiałów niemetalowych stosowane są czujniki pojemnościowe albo optyczne. W tych pierwszych,

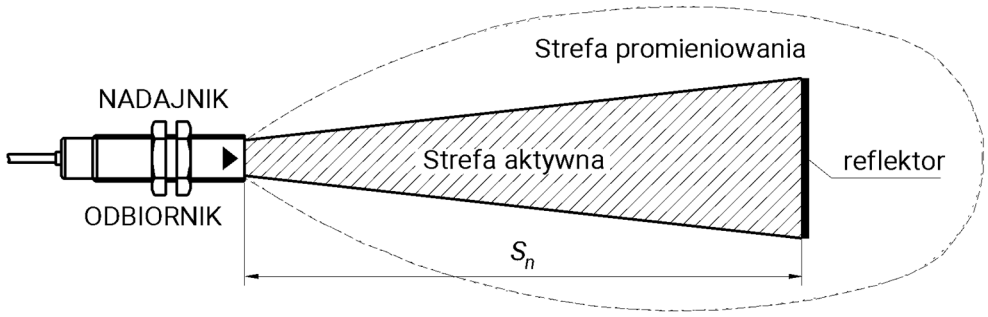
zamiast cewki indukcyjnej jest stosowany kondensator. Zbliżenie obiektu do czoła czujnika powoduje zmianę pojemności kondensatora, którego okładki znajdują się w „czole" czujnika. Zakres zmian pojemności zależy od przewodności wykrywanego materiału, jego stałej dielektrycznej, jak też jego masy, powierzchni i odległości od elektrody. Zmiana pojemności kondensatora, spowodowana zbliżeniem jakiegoś materiału/elementu powoduje wzrost napięcia generowanego przez oscylator RC w czujniku. Sygnał ten steruje wyjściem. Czujniki pojemnościowe to elementy automatyki reagujące na wprowadzenie w ich strefę czułości metali, szkła, tworzyw, drewna, materiałów sypkich, cieczy i różnych substancji. Ich częstotliwość pracy jest znacznie mniejsza od częstotliwości czujników indukcyjnych i wynosi zwykle do ok. 50 Hz.

Czujniki optoelektroniczne wykrywają obiekty, które znajdują się na drodze przebiegu wiązki światła. Wykorzystuje się je m.in. do kontroli położenia ruchomych części maszyn, do identyfikacji obiektów znajdujących się w zasięgu działania czujników, np. przesuwających się na taśmach transportowych, do określania poziomu cieczy i materiałów sypkich itp. Zaletą czujników optoelektronicznych jest duży zasięg działania uzyskiwany dla małych obudów czujników. Najczęściej czujniki te wykorzystują modulowane światło z zakresu bliskiej podczerwieni. Zaletą takiego rozwiązania jest niewrażliwość czujników na widzialne światło z otoczenia. Dodatkowo dzięki wzajemnej synchronizacji nadajnika i odbiornika, gwarantowana jest duża odporność na zakłócenia i możliwość pracy w warunkach zanieczyszczenia powietrza i przy zabrudzeniu układu optycznego czujnika. Wytworzony w nadajniku impuls świetlny, nawet osłabiony rozproszeniem, dociera do odbiornika lub nie, co jest sygnalizowane stanem 0 albo 1 na wyjściu.



Rys. 7. Jednowiązkowa bariera świetlna

Zaletą czujników optycznych jest duża strefa działania, która może sięgać nawet kilkudziesięciu metrów. Zanieczyszczenie powietrza i zabrudzenie układu optycznego skracają tą strefę. Jednym z najstarszych czujników optycznych jest jednowiązkowa bariera świetlna, w której nadajnik i odbiornik są umieszczone w oddzielnych obudowach (rys. 7). Stosuje się także bariery świetlne, w których występuje wiązka (ściana) światła, przebiegająca od nadajników do odbiorników, które znajdują się naprzeciw siebie w skrajnych punktach zasięgu. Przesłonięcie wiązki promieni świetlnych przez obiekt powoduje przerwanie transmisji fali świetlnej i przełączenie obwodu wyjściowego czujnika.



Rys. 8. Czujnik refleksyjny

W czujnikach optoelektronicznych refleksyjnych i odbiciowych nadajnik i odbiornik są umieszczone we wspólnej obudowie (rys. 8). Są one skierowane w końcowy punkt zasięgu, w którym jest umieszczony specjalny reflektor odblaskowy. Od tego reflektora odbija się wysyłana przez nadajnik wiązka promieni świetlnych. Jej przesłonięcie przez obiekt powoduje przerwanie transmisji i przełączenie obwodu wyjściowego czujnika z „0” na „1”. Czujniki odbiciowe, wykrywają światło odbite nie od reflektora a od przedmiotu, którego obecność wykrywają. Podobnie jak reflektor, przedmiot wykrywany musi mieć zdolność odbijania światła.

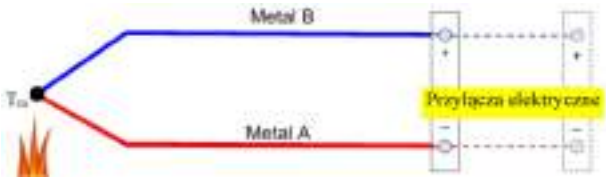
Czujniki ultradźwiękowe wykorzystują fale ultradźwiękowe (o częstotliwości powyżej 20 kHz), do wykrywania elementów albo do pomiaru ich odległości od czujnika. Jeśli w strefie działania czujnika jest jakiś obiekt, to wysyłane przez nadajnik impulsy dźwiękowe są od niego odbijane i wracają do odbiornika. Mierzony jest czas upływający od wysyłania impulsu ultradźwiękowego do powrotu echa odbitego od przeszkody. Odległość jest obliczana na podstawie tego czasu i prędkości dźwięku. Czujniki ultradźwiękowe pozwalają na

bezdotykowe wykrycie obiektów metalowych i niemetalowych. Są wykorzystywane głównie do pomiaru odległości obiektów od czujnika oraz do detekcji przedmiotu lub poziomu cieczy.

Czujniki pomiarowe

Osobną grupę elementów automatyki stanowią czujników pomiarowe. Zwykle muszą one współpracować z odpowiednim do nich układem elektronicznym, który formuje elektryczny sygnał wyjściowy, w taki sposób, aby jego wartość była proporcjonalna do wartości mierzonej. Mierzą one różne wielkości fizyczne, pozwalając na monitorowanie parametrów takich jak położenie, prędkość, siła, naprężenie ciśnienie, temperatura. Przetwarzają one wielkość fizyczną na sygnał elektryczny.

Przykładem czujnika analogowego jest termopara (rys. 9.) Temperatura musi być mierzona w wielu procesach technologicznych. W automatyce, celu dokonania pomiaru temperatury stosuje się różnego rodzaju przetworniki z wyjściem elektrycznym. Bardzo często do pomiaru temperatury wykorzystywane są termopary (termoogniwa). Są to czujniki temperatury wykorzystujące zjawisko Seebecka. Polega ono na powstawaniu zależnej od temperatury siły elektromotorycznej na styku dwóch różnych metali. Termopara składa się z dwóch różnych metali (drucików), spojonych na jednym końcu (punkt pomiaru).



Rys. 9. Termopara

Najczęściej stosowane są termopary typu:

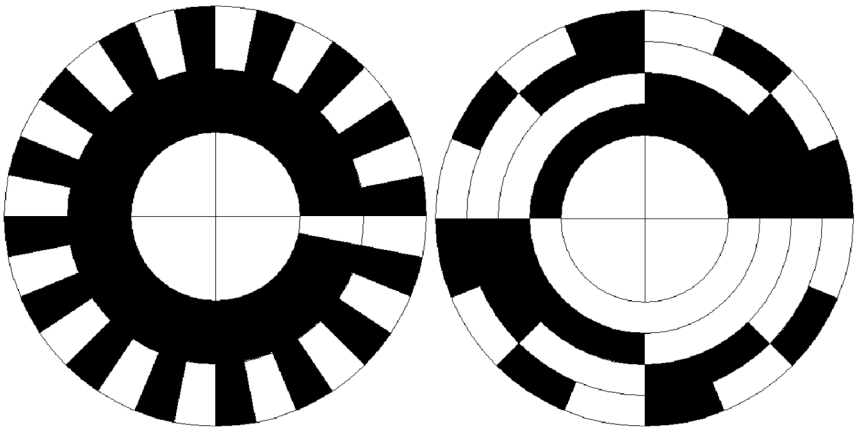
- T: miedź (Cu) — konstantan (Cu+Ni),
- J: żelazo (Fe) — konstantan (Cu+Ni),
- K: chromel (Ni+Cr) —alumel (Ni+Al),
- N: (Ni+Cr+Si) — (Ni+Si),
- E: chromel (Ni+Cr) — konstantan (Cu+Ni).

Przykładowo, termopara J może mierzyć temperatury od -210 do +1200 ºC, generując napięcie od -8,1 do 69,5 mV.

Termorezystory to czujniki, których rezystancja zmienia się wraz ze zmianą temperatury. Wielkość zmiany rezystancji określana jest przez temperaturowy współczynnik rezystancji. Termorezystory platynowe mogą mierzyć temperaturę w zakresie –200°C ÷ +850°C.

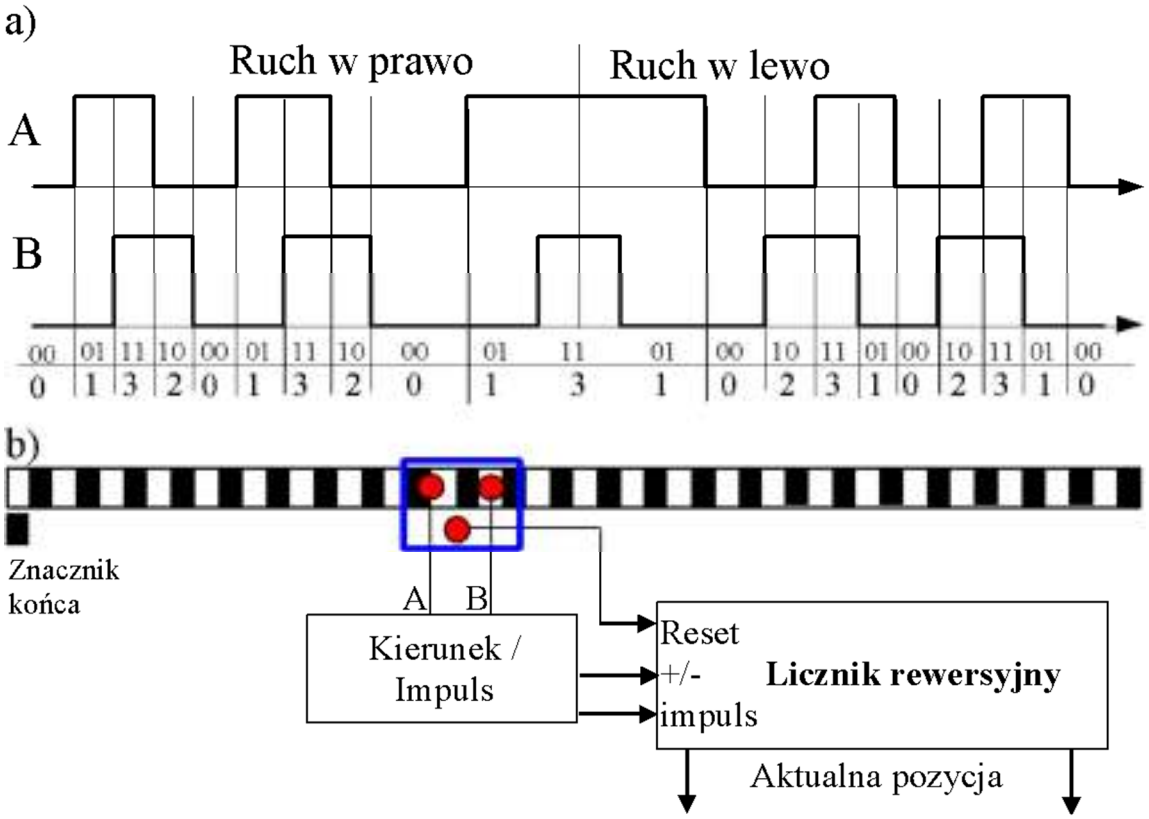
W pomiarach przemysłowych najczęściej wykorzystuje się czujniki platynowe Pt100 o rezystancji 100 Ohm w temperaturze 0 °C. W automatyce wykorzystuje się również termorezystory niklowe typu: Ni100, Ni1000, które mogą być wykorzystywany do pomiaru temperatury w zakresie –60 °C ÷ +180 °C. Termorezystory miedziane zastosowanie znajdują głównie w chłodnictwie. Ich zakres pomiarowy to 0 °C ÷ +150 °C. Drugim sposobem pomiaru temperatur są pomiary bezkontaktowe, przy pomocy pirometrów oraz kamer termowizyjnych. Pirometr mierzy głównie temperaturę punktu lub małej powierzchni elementu. Kamera termowizyjna rejestruje rozkład temperatur na powierzchni, tj. kilkaset tysięcy punktów w przestrzeni obejmowanej przez jej obiektyw. Pozwala na obrazowanie rozkładu przestrzennego pola temperatury na powierzchni obiektu.

a) b)



Rys. 10. Tarcze enkoderów: a) inkrementalnego,
b) absolutnego 4-ro bitowego

Do pomiaru pozycji i przemieszczenia bardzo powszechnie stosowanymi w systemach automatyki są enkodery (rys. 10). Budowane są wersje obrotowe i liniowe. Dzieli się je na przyrostowe (inkrementalne) i absolutne (kodowe). W trakcie ruchu, enkodery inkrementalne generują impulsy, które są zliczane przez specjalne liczniki rewersyjne tj. dodające-odejmujące. Enkoder pokazany na rysunku 10a generuje 16b impulsów na obrót. Enkodery absolutne generują liczby, określające aktualne położenie. Pokazany na rysunku, generuje na jeden obrót 16 liczb 4-ro bitowych.



Rys. 11. Enkoder inkrementalny: a) generowane impulsy, b) liniał enkodera podłączony do układu pomiarowego z licznikiem rewersyjnym

Enkodery liniowe wyposażone są w liniał, będący wzorcem długości (rys. 11a). Ma on naniesione w równych odległościach, na przemian ciemne i jasne pola albo szczeliny, przez które przechodzi wiązka nadawana do odbiornika. Wzdłuż liniału przesuwa się czytnik z fotodiodą, rozpoznający jasne i ciemne pola albo szczeliny. Od liczby szczelin zależy rozdzielczość urządzenia. Na tej podstawie odbiornik wysyła sygnały w kształcie impulsów 0 albo 1. Enkodery przyrostowe, inaczej inkrementalne, generują impuls co określone przesunięcie np. co 10 μm i dlatego muszą współpracować ze specjalnym licznikiem rewersyjnym, czyli liczącym w przód (dodającym) i w tył (odejmującym). Enkodery te wyposażone są w dwa czujniki generujące sygnały A i B przesunięte względem siebie (rys. 9b) o +90º albo o –90º. Na podstawie tego przesunięcia układ licznika wykrywa czy ruch odbywa się w lewo i wtedy dodaje impulsy, czy też w prawo i wtedy odejmuje impulsy (rys. 11c). Zmiana stanu licznika dokonywana jest przy zmianie dowolnego zbocza dla obu sygnałów A i B. Po włączeniu zasilania konieczne jest przesuniecie głowicy pomiarowej czujnika do

punktu odniesienia, tj. krańcowego, co zeruje licznik. Stanowi to istotną niedogodność enkodera inkrementalnego.

Rozwiązaniem konieczności zerowania układu pomiarowego po włączeniu zasilania jest zastosowanie enkodera absolutnego, w którym każdemu położeniu przyporządkowana jest konkretna wartość liczbowa, określająca aktualną pozycję (rys. 10). Dzięki temu enkoder absolutny podaje zaraz po włączeniu zasilania, na swoim wyjściu aktualną pozycję. Zanik napięcia nie powoduje utraty informacji o aktualnym położeniu i nie ma konieczności ponownego ustalania punktu zerowego, według którego będą określane przemieszczenia. Enkoredy wykonywane są w wersji obrotowej i liniowej.

Enkodery absolutne są droższe od inkrementalnych o takich samych parametrach. Na rynku można także spotkać na Enkodery optyczne i magnetyczne. Enkodery magnetyczne obrotowe mają pierścień i czujnik, który mierzy przemagnesowania, natomiast magnetyczne enkodery liniowe z elementu pomiarowego poruszającego się nad taśmą magnetyczną. Są one nazywane liniałami magnetycznymi. Enkodery liniowe są stosowane w maszynach CNC, centrach obróbczych i obrabiarkach elektroiskrowych. W najdokładniejszych stosuje się interferencyjne układy pomiarowe, w których prążki (wzorce długości) uzyskuje się w wyniku interferencji przenikających się wiązek lasera. Dzięki interpolacji można uzyskać rozdzielczości nawet do 0,01 μm. W Tabeli 1 zestawiono różne, najważniejsze czujniki.

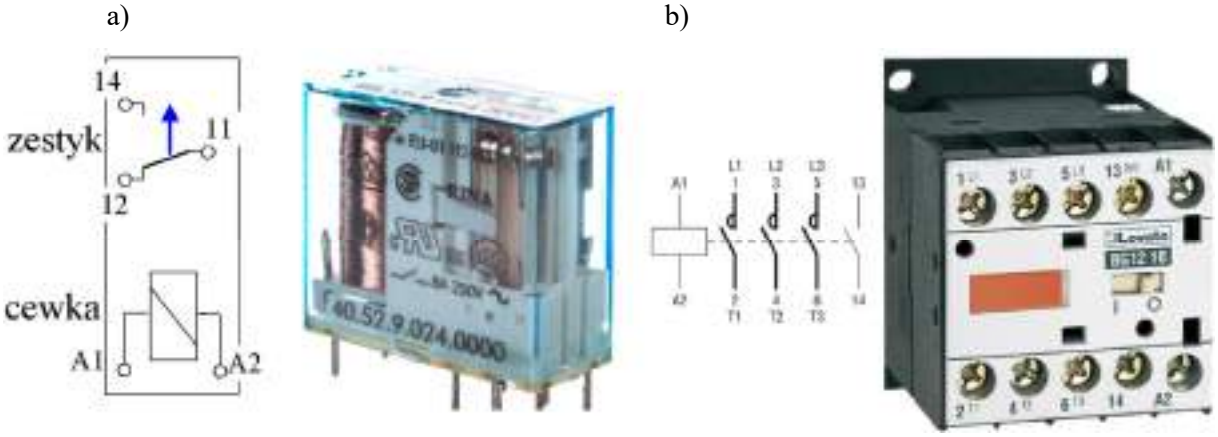
Tabela. 1 Klasyfikacja czujników

Wielkość mierzona	**Typ czujnika**
Położenie kątowe	Potencjometryczne, transformatory położenia kątowego, enkodery
Przemieszczenie liniowe	Potencjometryczne, indukcyjne różnicowe LVDT, magnetostrykcyjne
	Enkodery inkrementalne i absolutne: optyczne, inukcyjnościowe, pojemnościowe
	Enkodery inkrementalne optyczne i magnetyczne
Prędkość i przyspieszenie	Prądnice: tachometryczne, synchroniczne, indukcyjne, optyczne
	Akcelerometry, żyroskopy
Siła	Tensometry rezystancyjne i półprzewodnikowe, piezoelektryczne Mostki tensometryczne
Dotyk	Stykowe: mikroprzełączniki, krańcówki
	Pojemnościowe, rezystancyjne
Wykrywanie obecności	Ultradźwiękowe, indukcyjnościowe, optyczne, pojemnościowe, stykowe
	Optyczne: diodowe i laserowe: bariera, odbiciowe, refleksyjne
	Systemy wizyjne z kamerami 2D i 3D

Identyfikacja obiektu	Czytniki kodów paskowych i RFID, systemy wizyjne
	Termopary, termorezystory
Temperatura	Termistory półprzewodnikowe, pirometry, diody PIR
	Kamery termowizyjne

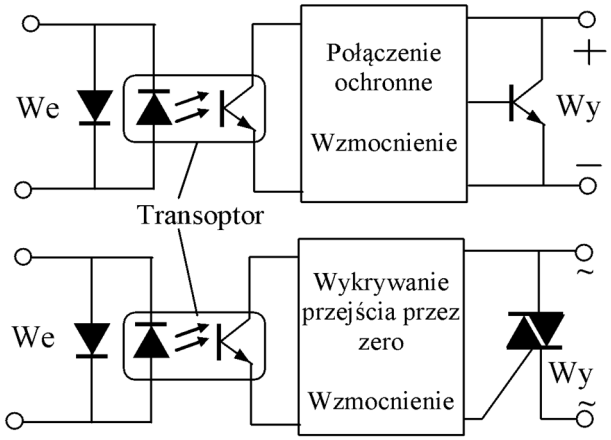
3. Napędy i elementy wykonawcze

Przekaźniki i styczniki są elementami elektromechanicznymi, wykonawczymi stosowanymi powszechnie w systemach automatyki do włączania i wyłączania urządzeń. Maksymalna częstotliwość ich przełączania mieści się w zakresie 0,5÷5 Hz. Bardzo małe przekaźniki mogą być przełączane sygnałem o częstotliwości nawet do 70 Hz, jednak charakteryzują się oneto ich niewielka trwałość, ograniczenia w stosowaniu oraz cena powodują, że są one rzadko stosowane. Sygnałem wejściowym przekaźnika jest napięcie i natężenie prądu elektrycznego podawanego na cewkę, a sygnałem wyjściowym – stan styków (zamknięte lub otwarte). Wyróżniamy przekaźniki, których cewki są zasilane napięciem stałym albo przemiennym. W katalogach podaje się następujące podstawowe parametry przekaźników: napięcie znamionowe cewki (np. 6; 12; 24 VDC albo 24VAC, 230VAC), moc znamionową cewki, liczbę i rodzaj styków, moc łączeniową przy obciążeniu rezystancyjnym, maksymalną częstotliwość przełączania oraz trwałość łączeniową i mechaniczną. Bardzo istotnym parametrem jest właśnie trwałość przekaźnika elektromechanicznego. Minimalna trwałość mechanizmu wynosi od 1 do 100 mln cykli, a trwałość styków zależy od rodzaju obwodu, który jest sterowany przekaźnikiem. Można ją zwiększyć tylko przez prawidłowe dobranie typu przekaźnika do rodzaju obciążenia. Większość rzeczywistych obciążeń ma charakter pojemnościowy, co oznacza, że prąd pobierany w pierwszej chwili po zamknięciu obwodu jest znacznie większy niż kilka chwil później.



Rys. 12. Symbol i widok: a) przekaźnika, b) stycznika 3-fazowego

Jeśli przez styki płynie natężenie prądu bliskie maksymalnemu, to dla zapewnienia wysokiej żywotności, napięcie przełączane nie powinno przekraczać 10÷20% wartości maksymalnej. Na trwałości styków oraz ich prądowo-napięciową charakterystykę duży wpływ mają materiały wykorzystywane na pokrycie pól kontaktowych. Renomowani producenci produkują przekaźniki ze stykami wykonanymi z materiałów, które są przystosowane do określonych warunków fizycznych oraz rodzaju zasilanego odbiornika. Rodzaj kategorii użytkowania stycznika lub przekaźnika jest ważnym parametrem, a prawidłowy jego dobór sprzyja długiej i bezawaryjnej pracy.



Rys. 13. Przekaźniki półprzewodnikowe: tranzystorowe i triakowe

Styczniki to łączniki mechaniczne, które działają na tej samej zasadzie co przekaźniki elektromagnetyczne. Są też podobne pod względem budowy, ale znacznie większe. Różnica dotycząca typu łączonych za ich pomocą obwodów określa ich zastosowanie. Styczniki służą do włączania i wyłączania obwodów głównych elektrycznych dużej mocy, typu grzałki, silniki i elektromagnesy, a przekaźniki do przełączania obwodów pomocniczych. Styczniki mają zwykle trzy zestyki główne dużej mocy do zasilania obwodów trójfazowych np. 3 x 600VAC/50A oraz zestyki pomocnicze (np. do podtrzymania). Ich cewki są najczęściej przystosowane do zasilania napięciem 230 VAC oraz 24VDC. Częstotliwość przełączania styczników jest niewielka - mogą załączać obwody tylko 0,5÷2 razy na sekundę.

Na rynku dostępne są także przekaźniki półprzewodnikowe (ang. *solid state relays* – SSR). Stanowią one elektroniczny odpowiednik przekaźników elektromechanicznych. Zamiast cewki stosowane w nich są transoptory, składające się z diody świecącej i fototranzystora (rys. 13). Zapewniają one izolację między układami o różnicy potencjałów nawet do 4 kV, chroniąc

dodatkowo wrażliwe elementy elektroniczne. Na wyjściu jest używany tranzystor dużej mocy, tyrystor albo triak. Tym samym przełączanie poszczególnych obwodów realizowane jest bez udziału części ruchomych, czyli bez ruchomych styków. Przekaźniki wykonują sterowanie bezstykowe lub elektroniczne. Wszystkie przekaźniki i styczniki zapewniają oddzielenie elektryczne obwodów automatyki od obwodów dużej mocy, zapewniając tzw. separację galwaniczną. Przekaźniki SSR charakteryzują się następującymi cechami:

- dużą częstotliwością przełączania, dochodzącą do kilku kHz,
- bezstykową, bezgłośną oraz iskrobezpieczną pracą,
- bardzo dużą trwałością i niezawodnością,
- zdolnością do załączania dużych mocy (prąd od 90 do 150 A i napięcie do 1000 V),
- sterowaniem chwili włączenia i wyłączenia np. przy przechodzeniu sinusoidy przez zero.

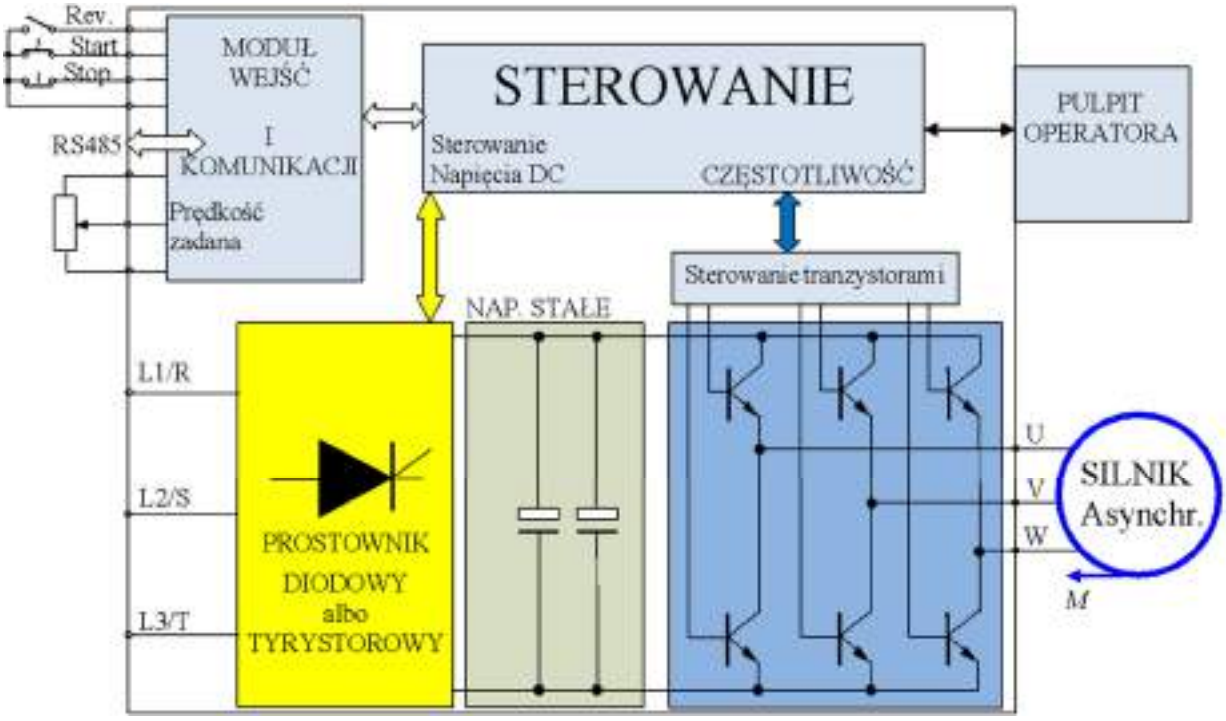
Zależnie od sposobu i rodzaju dostarczanej energii wyróżniamy urządzenia napędowe:

- pneumatyczne (brak zagrożenia pożarowego, odporność na wpływ płynów i związków agresywnych, łatwość w eksploatacji, niezawodność w działaniu, nie wymagają stosowania zabezpieczeń przed przeciążeniem),
- elektryczne (nośnikiem jest sygnał elektryczny – najczęściej napięcie lub prąd stały, bardzo bogata i rozbudowana dostępność urządzeń pomiarowych i sterujących, łatwość wykonania operacji matematycznych, przesyłanie sygnałów na dowolne odległości),
- hydrauliczne (możliwość uzyskania bardzo dużych sił przy dobrej dynamice, duża trwałość, bardzo dobry stosunek energii do masy).

Do napędzania urządzeń montowanych na automatyzowanych liniach produkcyjnych najczęściej stosuje się silniki elektryczne. Do wykonywania prostych ruchów liniowych od jednej pozycji krańcowej do drugiej najczęściej stosowane są siłowniki pneumatyczne, sterowane zaworami elektropneumatycznymi. Silniki wirujące dzieli się na dwie zasadnicze grupy: prądu stałego i prądu przemiennego. Silnik prądu stałego jest silnikiem elektrycznym zasilanym prądem stałym i służy do zamiany energii elektrycznej na energię mechaniczną. Jako maszyna elektryczna prądu stałego może pracować jako silnik lub prądnica. W tym drugim przypadku wirnik napędzany jest energią mechaniczną dostarczoną z zewnątrz, a na zaciskach uzwojenia twornika odbierana jest wytworzona energia elektryczna. Większość silników prądu stałego to silniki komutatorowe, to znaczy takie, w których uzwojenie twornika zasilane jest prądem poprzez komutator. Jednak istnieje wiele silników prądu stałego które nie posiadają komutatora lub też komutacja przebiega na drodze elektronicznej.

Maszyny prądu przemiennego buduje się jako trójfazowe i jednofazowe. Ze względu na stosunkowo prostą ich budowę maszyny prądu przemiennego znalazły powszechne zastosowanie jako silniki napędzające różnego rodzaju urządzenia mechaniczne oraz jako generatory wielkiej mocy. Maszyny prądu stałego mają wprawdzie budowę bardziej skomplikowaną od maszyn prądu przemiennego, lecz równocześnie odznaczają się lepszymi właściwościami regulacyjnymi. Dlatego też są stosowane w bardziej skomplikowanych, wymagających np. regulacji prędkości obrotowej, układach napędowych.

Napędy z silnikami asynchronicznymi trójfazowymi



Rys. 14. Schemat blokowy przemiennika częstotliwości – falownika

Obecnie najczęściej stosowanymi w maszynach i urządzeniach technologicznych są silniki trójfazowe indukcyjne. Na ich stojanie nawinięte są trzy albo sześć uzwojeń tworzących odpowiednio, jedną albo dwie pary biegunów. Są zasilane napięciem trójfazowym, które jest źródłem wirującego pola magnetycznego o prędkości synchronicznej 50 Hz, a prędkość wirowania wynosi 1500 obr/min. Wirnik silnika asynchronicznego jest najczęściej wykonany w postaci szeregu prętów zwartych po obu końcach (tzw. silniki klatkowe). Wirujące, czyli zmienne pole magnetyczne powoduje indukowanie siły elektromotorycznej (SEM) w prętach albo uzwojeniach wirnika. Wartość tej siły jest proporcjonalna do różnicy prędkości wirującego

pola i wirnika. W rezultacie przez zwarte uzwojenia wirnika płynie prąd wytwarzający pole magnetyczne. Wirujące pole magnetyczne stojana przyciąga i wprawia w ruch obrotowy wirnik, który generuje moment obrotowy na wale wyjściowym. Wirnik obraca się z prędkością mniejszą od prędkości wirowania pola stojana. W rezultacie występuje tzw. poślizg, czyli opóźnianie się wirnika w stosunku do pola stojana, W trakcie rozruchu, tj. przy zatrzymanym wirniku silnik rozwija maksymalny moment zwany rozruchowym M_r. W trakcie pracy poślizg zmienia się od bliskiego 1% do ok. 15%. Zwiększenie momentu obciążenia powyżej wartości M_k prowadzi do zatrzymania wirnika.

Prędkość obrotową można regulować przez zmianę prędkości wirującego pola, tzn. przez zmianę częstotliwości napięcia zasilającego *f*. Dlatego aby wytworzyć w stojanie pole magnetyczne wirujące o różnej częstotliwości, stosuje się przemienniki częstotliwości, nazywane też falownikami. Są to zasilacze zmiennoprądowe z regulowaną bezstopniowo częstotliwością napięcia zasilającego silnik. Zmiana częstotliwości *f* napięcia zasilającego powoduje nie tylko zmianę prędkości silnika, lecz również zmianę momentu elektromagnetycznego rozwijanego przez silnik. Zmniejszanie częstotliwości (poniżej 50 Hz) powoduje wzrost momentu elektromagnetycznego, a zwiększenie częstotliwości prowadzi do zmniejszenia momentu elektromagnetycznego. Dlatego, dla poprawnej pracy silnika, niezależnie od aktualnej częstotliwości zasilania, konieczne jest spełnienie warunku: U/f = const. Przemiennik częstotliwości, nazywany też falownikiem, zawiera 4 bloki funkcjonalne:

– układ sterowania i komunikacji,
– prostownik trójfazowy, najczęściej regulowany,
– pośredni obwód filtracji napięcia stałego z baterią kondensatorów,
– blok tranzystorowy przekształcający prąd stały w trójfazowy prąd przemienny o częstotliwości wynikającej z zadanej prędkości obrotowej.

Schemat blokowy falownika został pokazany na rys. 17. Napięcie przemienne 3 x 400VAC o częstotliwości 50Hz z sieci zasilającej jest prostowane do napięcia stałego oraz filtrowane (wygładzane) za pomocą kondensatorów. Może być sterowany albo nie. Zaletą prostownika sterowanego jest możliwość regulowania napięcia DC przez tyrystory. Stopień pośredni w przetwornicy, to bateria kondensatorów. Można ją traktować jako magazyn energii, z którego układ sterowanych tranzystorów kształtuje trójfazowe napięcia o regulowanej częstotliwości i wartości. Tranzystory pracują jak klucze, tzn. otwierają lub zamykają przepływ prądu ze źródła napięcia stałego do odpowiedniej fazy silnika. Układ sterowania otrzymuje

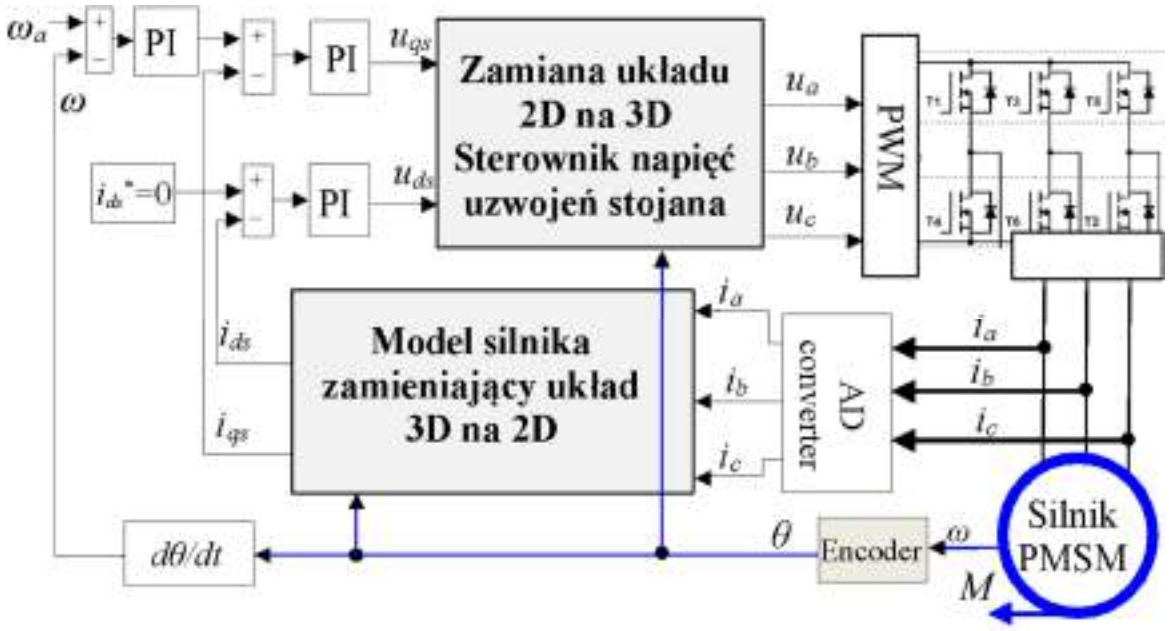
i obsługuje sygnały komunikacyjne z otoczenia przetwornicy oraz steruje pracą inwertera mocy. Sygnały te mogą pochodzić z zewnętrznych urządzeń sterujących bądź z panelu operatora.

Napędy z silnikami synchronicznymi PMSM

Spośród silników synchronicznych w urządzeniach technologicznych praktyczne zastosowanie w układach napędowych znalazły silniki synchroniczne z magnesami trwałymi umieszczonymi na wirniku. Na ich stojanie nawinięte jest trójfazowe uzwojenie. Takie silniki nazywane są też Permanent Magnet Synchronous Motor – PMSM. Są stosowane głównie w napędach posuwowych obrabiarek i w robotach. Silniki elektryczne z magnesami trwałymi umieszczonymi na wirniku mają najwyższą sprawność energetyczną ze wszystkich znanych i stosowanych rodzajów maszyn elektrycznych. Stojan posiada trójfazowe uzwojenie, wytwarzające kołowe pole wirujące. Rozruch i sterowanie w warunkach zmian obciążenia musi być kontrolowane elektronicznie. Silniki te mogą pracować, w zależności od sposobu zasilania i sterowania, jako silniki:

- synchroniczne PMSM,
- bezszczotkowe prądu stałego z komutatorem elektronicznym sterowane skokowo.

W obu przypadkach właściwości napędu są inne, decyduje o tym rozkład pola magnetycznego w szczelinie silnika oraz sposób zasilania i sterowania. Silniki PMSM powinny być sterowane napięciami sinusoidalnymi. Do tego konieczne jest stosowanie dokładnego pomiaru położenia wirnika za pomocą np. enkodera. Informacja o położeniu wirnika dostarczana jest do sterownika, który odpowiednio steruje zespołem mocy falownika. Dodatkowo konieczny jest pomiar natężenia prądu w poszczególnych uzwojeniach. Silnik synchroniczny z magnesami trwałymi sterowany jest z falownika metodą PWM, kluczami tranzystorowymi, na podstawie pomiaru prądów poszczególnych faz silnika oraz sygnału z enkodera (rys. 15).



Rys. 15. Układ sterowania silnikiem PMSM

W obu przypadkach właściwości napędu są inne, decyduje o tym rozkład pola magnetycznego w szczelinie silnika oraz sposób zasilania i sterowania. Silniki PMSM powinny być sterowane napięciami sinusoidalnymi. Do tego konieczne jest stosowanie dokładnego pomiaru położenia wirnika za pomocą np. enkodera. Informacja o położeniu wirnika dostarczana jest do sterownika, który odpowiednio steruje zespołem mocy falownika. Dodatkowo konieczny jest pomiar natężenia prądu w poszczególnych uzwojeniach. Silnik synchroniczny z magnesami trwałymi sterowany jest z falownika metodą PWM, kluczami tranzystorowymi, na podstawie pomiaru prądów poszczególnych faz silnika oraz sygnału z enkodera (rys. 15).

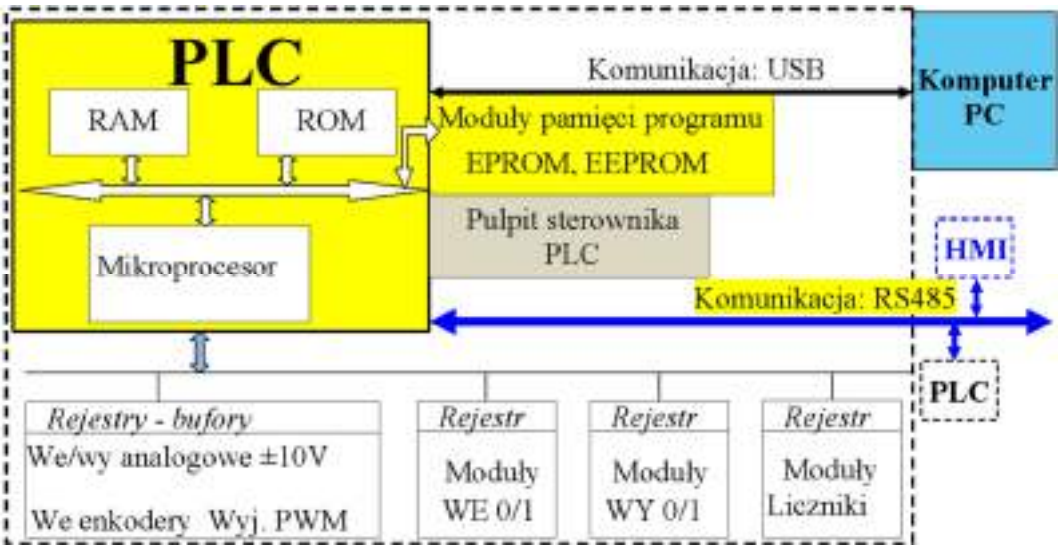
Przeważająca liczba zadań sterowania urządzeniami w praktyce przemysłowej sprowadza się do załączania i wyłączania wybranych elementów napędowych, grzałek, elektromagnesów i lamp. Nazywa się to sterowaniem dwustanowym, przełączającym lub binarnym. Projektowanie takiego sterowania sprowadza się do zbudowania schematu urządzenia sterującego z wykorzystaniem przełączników i przekaźników albo bramek logicznych. Schemat połączeń tych elementów stanowi algorytm sterowania, którego techniczna realizacja polega na wprowadzeniu tego schematu do pamięci uniwersalnego sterownika. Jest on zdolny do wykonywania algorytmu symulując niejako wprowadzony schemat. Mówimy wtedy o sterowniku programowanym pamięciowo. Program to ciąg kolejno następujących i logicznie powiązanych ze sobą instrukcji do przetwarzania sygnałów i danych przez urządzenie sterujące - sterownik.

4. Podstawy sterowników przemysłowych typu PLC

Do dyspozycji automatyków jest dzisiaj bardzo różnorodna aparatura pomiarowa tj. czujniki i systemy pomiarowe, silniki i kompletne napędy, regulatory, sterowniki, panele do wprowadzania i wyświetlania informacji oraz systemy komunikacji.

Sterowniki programowalne (ang. *Programable Logic Controllels* – PLC), nazywane także binarnymi, są obecnie powszechnie stosowane w przemyśle. Na podstawie sygnałów analogowych i cyfrowych o stanie urządzenia, otrzymywanych z czujników oraz urządzeń pomiarowych, sterowniki PLC przetwarzają dane zgodnie z zapisanym w pamięci programem i generują sygnały sterujące. Pierwsze sterowniki PLC zbudowano w końcu lat 60. XX w. Ich główną zaletą jest możliwość szybkiej zmiany wykonywanego algorytmu przez programowanie pamięci sterownika. Charakteryzują się one dużą niezawodnością w warunkach przemysłowych, przy małych wymiarach oraz łatwością utrzymania w ruchu produkcyjnym, dzięki możliwości napraw przez wymianę całych modułów. Na początku lat 80. XX w. pojawiły się inteligentne moduły we/wy, mające własne procesory, które mogły realizować złożone funkcje obliczeniowe. Od lat 90. XX w. wykorzystuje się komputery PC do programowania sterowników PLC, co znacznie rozszerzyło możliwości programowe i komunikacyjne sterowników, pracujących obecnie w sieciach.

Na rysunku 16 pokazano schemat blokowy typowego sterownika PLC. Jego głównym elementem jest mikroprocesorowa jednostka centralna (CPU), która wykonuje obliczenia logiczne i arytmetyczne oraz zarządza sterownikiem. Jest wyposażona w pamięć typu ROM (ang. *Read Only Memory*) przeznaczoną do odczytu. Są w niej zapisane informacje o konfiguracji sterownika. W pamięci programowanej (EEPRM) zapisywany jest program sterujący, napisany przez użytkownika. W pamięci RAM (ang. *Random Acces Memory*) są przechowywane dane odczytywane z wejść oraz wyniki wykonywanych operacji, w tym te określające stany wyjść. Pamięć RAM jest zwykle wykorzystywana do budowy elementów czasowych i licznikowych. Do podłączenia czujników binarnych, sterowniki PLC są wyposażone w moduły wejść 0/1. Podobnie do podłączania elementów wykonawczych, które można tylko włączyć albo wyłączyć, stosowane są przekaźnikowe moduły wyjściowe. Poza tym, sterowniki wyposażone są także w liczniki, które mogą służyć do zliczania impulsów podawanych na wejścia albo do odmierzania czasu (Timery).

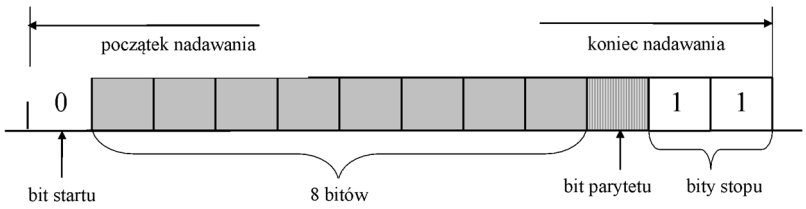


Rys. 16. Schemat blokowy sterownika przemysłowego PLC

Sterowniki PLC były początkowo przeznaczone tylko do wykonywania jednobitowych operacji przełączających i dlatego posiadały tylko wejścia binarne (0/1) na napięcie 0V/24VDC oraz wyjścia binarne – przekaźnikowe, mogące włączać elementy o napięcie max. 250V i natężeniu prądu do 1A. Obecnie są dostępne także wyjścia tranzystorowe (24VDC) bądź triakowe (230VAC). Współczesne sterowniki PLC są także wyposażone w moduły wejść i wyjść analogowych, napięciowych (–10V… +10V). Do sterowników można także dołączyć enkodery inkrementalne – do wejść licznikowych. Moduły wyjściowe umożliwiają sterowanie silnikami krokowymi oraz układami mocy z modulacją szerokości impulsów (PWM) itd. Realizacja zadań sterowania rozbudowanymi zestawami maszyn wymaga połączenia sterowników PLC w sieć np. RS485. Opcjonalnie do sterowników można też podłączyć monitory i panele operatorskie (ang. *Human-Machine Interface – HMI*), wykorzystywane do wizualizacji procesu i sterowania nim w trybie bezpośrednim (ang. *on-line*). Pozwalają one na wybór odpowiednich opcji sterowania, wprowadzanie nastaw regulatorów, kontrolę i korektę ewentualnych błędów sterowania. Mają one duże możliwości i dlatego pozwalają (w zależności od użytego oprogramowania) zarówno na wizualizację i kontrolę procesu, jak i na edycję, wprowadzanie i testowanie samego programu sterującego. Komputer PC podłączany jest do sterownika na czas programowania.

Ponieważ te sterowniki są obecnie budowane na bazie komputerów, które przetwarzają tylko sygnały cyfrowe – liczby binarne, to napięcie elektryczne zostaje przetworzone na liczby w przetwornikach analogowo-cyfrowych (AC) na liczby. Przetwarzanie przebiega w dwóch fazach: próbkowanie i kwantowanie. W pierwszej napięcie wejściowe zostaje zapamiętane,

a w drugiej przetworzone na liczbę zapisaną w kodzie binarnym. Przetwornika AC zamienia sygnał analogowy (napięcie) na ciąg liczb L1, L2, L3 … podawanych na wyjście co określony czas. Jego podstawowe parametry to: zakres napięć wejściowych np. 0 do 1V, 0 do 2,5V; liczba bitów na wyjściu np. 8, 12, 16 oraz czas przetwarzania (wyzwalania) *T*, zwykle of 1 μs do 1 ms. Przetwornik 8-mio bitowy generuje na wyjściu liczby z zakresu od 0 do 255 a 12-to bitowy od 0 do 4095. Przetwornik cyfrowo-analogowy (CA), przetwarza w kierunku odwrotnym tzn. zamienia liczby binarne na odpowiadające im napięcia wyjściowe. Jednak sygnał analogowy w postaci napięcia może być przesyłany tylko na niewielkie odległości, bowiem jest podatny za zakłócenia. Na większe odległości można przesyłać sygnał prądowy o zakresie prądów 4 mA 20 mA, który jest znacznie bardziej odporny na zakłócenia. Jednak najlepszy do transmisji na duże odległości jest sygnał cyfrowy, pozwalający na przesyłanie dowolnych informacji i danych.



Rys. 17. Przebieg transmisji szeregowej 1-go słowa

W układach do komunikacji sieciowej kolejne liczby są przesyłane za pomocą transmisji szeregowej. Jest to rodzaj cyfrowej transmisji danych, w którym dane są przesyłane jednym przewodem (albo jedną parą). Poszczególne bity słowa (liczb) są przesyłane kolejno, bit po bicie. Stosowane są następujące interfejsy: RS-232, RS-422A, RS-485, I2C, SPI, USB, Ethernet, USB. W praktyce stosowana jest też łączność bezprzewodowa. Używane są 3 główne standardy: WiFi (60%), ZigBee (20%) oraz Bluetooth (20%). Dane są poprzedzone bitem startu (stan logiczny: 0), który jest sygnałem do rozpoczęcia transmisji (rys. 17). Po nim są przesyłane bity danych, od najmłodszego do najstarszego, potem bit parytetu (parzystości) a po nim następuje odstęp, nazywany bitem albo bitami stopu przed następnym znakiem (stan logiczny: 1). Stop może on być dłuższy, bo nie ma ograniczenia czasowego.

Na rysunku 18 pokazano podłączenie różnych elementów do wejść binarnych (przełączniki stykowe, krańcówki, czujniki bezkontaktowe itp.) oraz do wyjść binarnych (lampki, przekaźniki i styczniki, elektromagnesy oraz elektrozawory). Przekaźniki i styczniki przeznaczone są do włączania i wyłączanie elementów dużej mocy typu: grzałki, silniki i napędy). Ze względu na rodzaj wyróżnić można sygnały:

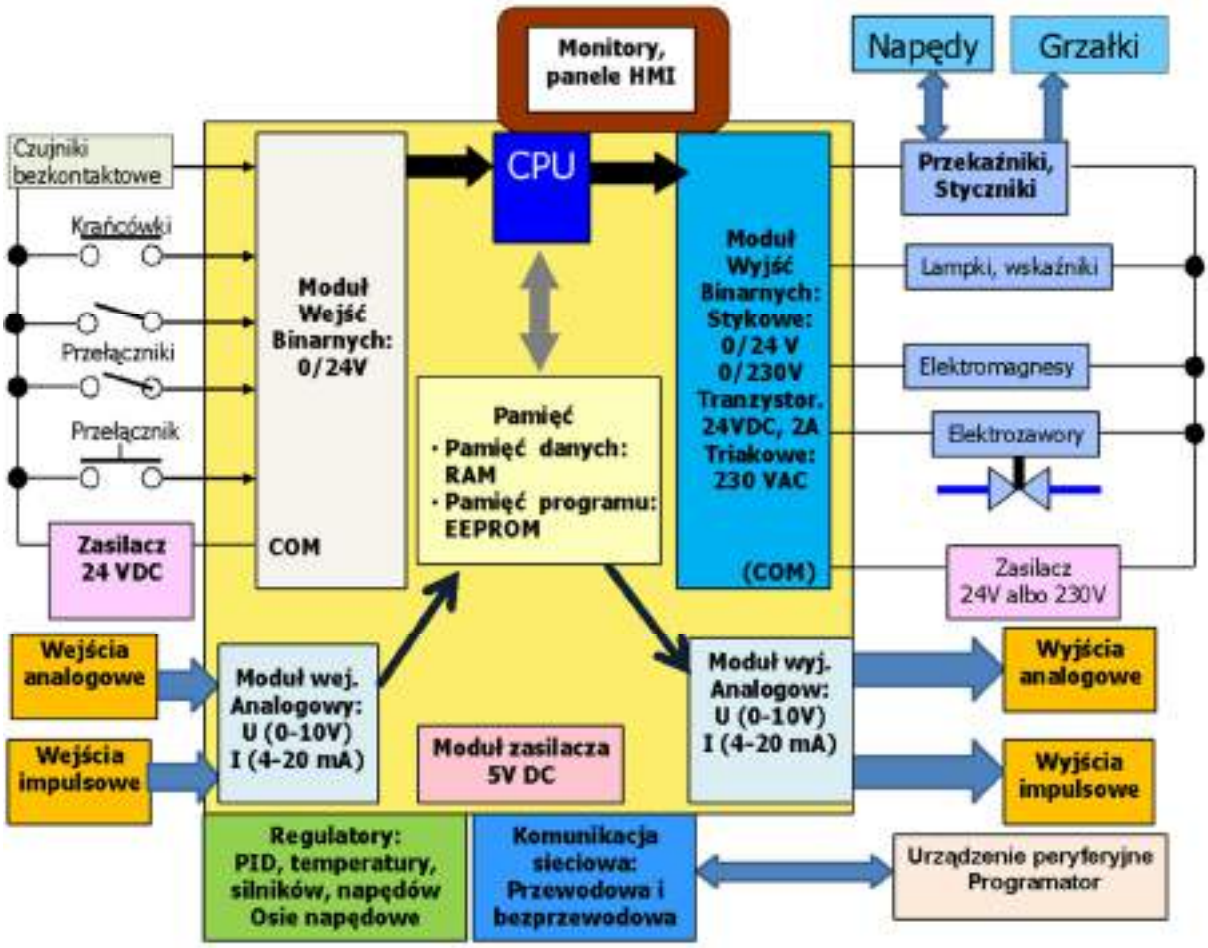
- binarne, przełączające, w których sygnał może przyjmować tylko jedną z dwóch wartości 0 lub 1, kodowaną napięciowo 0V/24VDC albo 0/230VAC i 0V/3x400VAC,
- analogowe albo ciągłe, w których sygnał może przyjmować dowolną wartość z określonego przydziału napięcia: –10V do +10V, 0 do 10V, 0 do 5V, albo natężenia prądu 0 do 20 mA albo 4 do 20 mA. Mogą być podłączone urządzenia pomiarowe, np. potencjometry, liniały transformatorowe, wagi tensometryczne natomiast do wyjść można podłączyć np. sterowniki napędów elektrohydraulicznych albo elektrycznych

Na wejścia sterownika PLC można podawać sygnały:

- binarne, które dostarczają informacje: on/off (24VDC/0VDC) z takich elementów jak:
  - wyłączniki krańcowe stykowe,
  - czujniki indukcyjne, pojemnościowe, optyczne itp.
  - przełączniki graniczne poziomu, ciśnienia itp.
- pomiarowe (analogowe), które dostarczają informacje ciągłe o aktualnych:
  - położeniu, prędkości, sile,
  - ciśnieniu, naprężeniu,
  - temperaturze, napięciu, natężeniu
- impulsowe w postaci upływu czasu *T* oraz o częstotliwości *f* określających:
  - zmiany położenia z enkoderów inkrementalnych
  - czas jaki upływa od wysłania impulsu do jego powrotu.

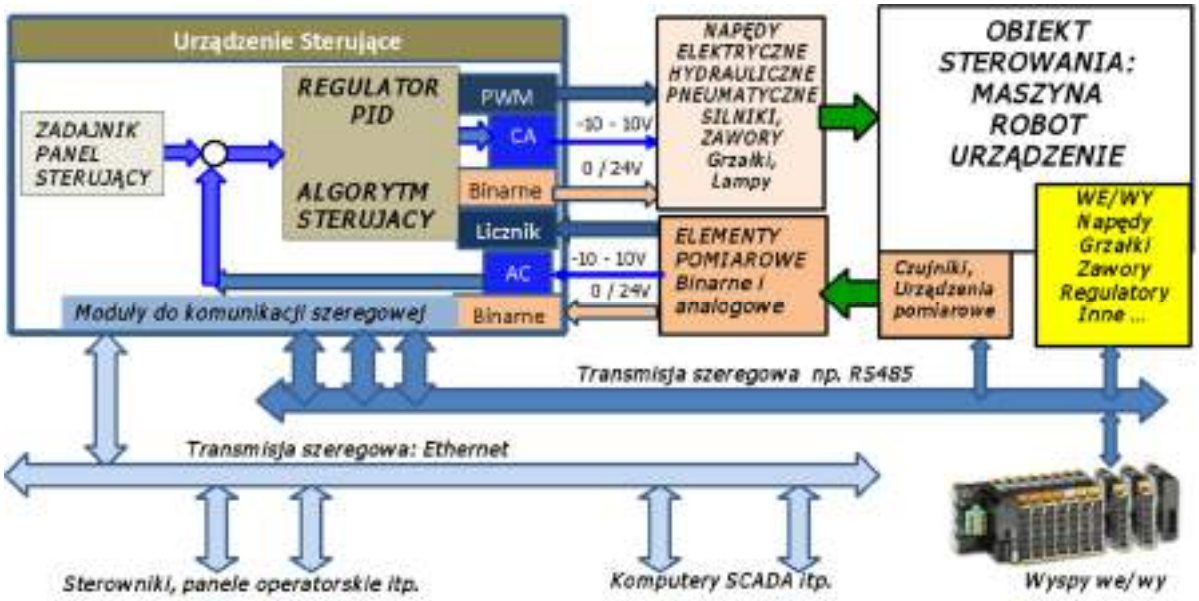
Wyjścia sterownika mogą być następujące:

- binarne typu on/off (0VDC/24VDC):
  - do włączania albo wyłączenia różnych elementów i urządzeń,
  - zwolnienie (enable), hamuj, kierunek L/P
- Analogowe do określenia zadanych (np. –10 do +10V):
  - generowanej siły, prędkości, położenia
  - ciśnienienia, natężenia przepływu, temperatury, napięcia itp.
- Impulsowe, czasowe do wykonywania kroków napędu albo jego prędkości (PWM).



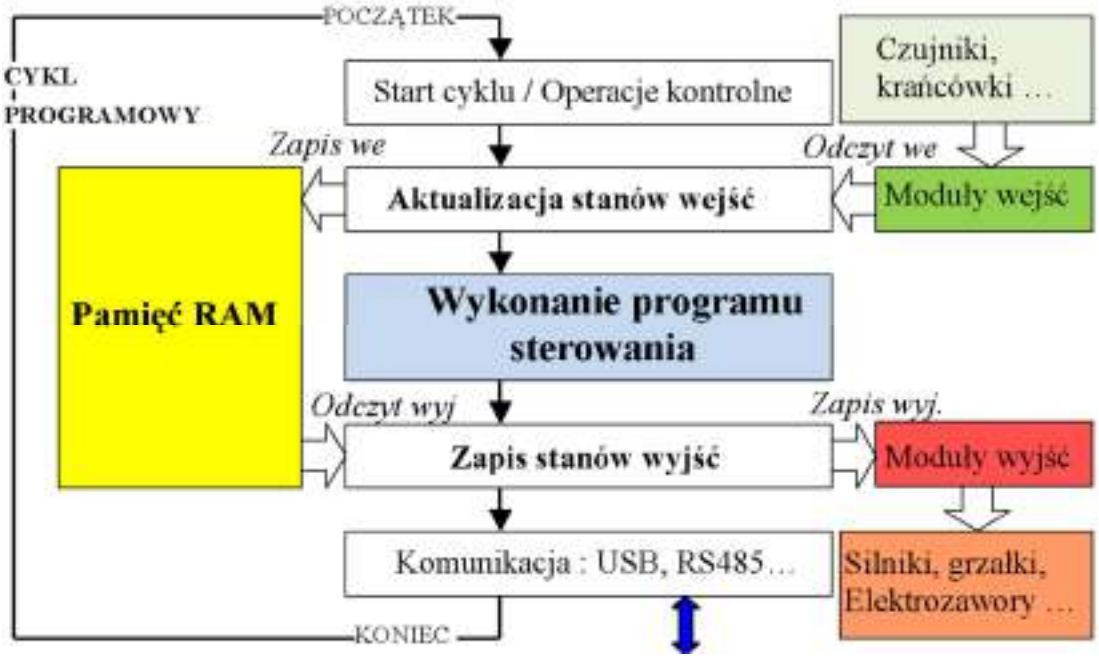
Rys. 18. Podłączenie elementów do sterownika przemysłowego PLC

Współcześnie produkowane sterowniki PLC wyposażone są także odrębne moduły różnych regulatorów np. PID albo regulatory dwupołożeniowe, przeznaczone do regulacji temperatury oraz regulatory silników i napędów. Dostępne są również moduły do regulacji i pozycjonowania od 1 do 4 napędów, tzw. regulatory kilku osi np. obrabiarek albo robotów. Sterowniki PLC mogą również komunikować się z różnymi urządzeniami pomiarowymi oraz napędowymi a także z innymi sterownikami PLC oraz komputerami PC, a tym samym z całym Internetem, za pośrednictwem łączności przewodowej bądź bezprzewodowej. Komunikacja ta jest prowadzona w sposób szeregowy, tzn. poszczególne słowa są przesyłane bit po bicie.



Rys. 19. Komunikacja elementów i układów automatyki

Do komunikacji systemu sterowania wykorzystywane są już od kilkunastu lat również magistrale szeregowe. Wśród nich, bardzo popularny jest standard RS485. Składa się z różnicowego nadajnika, dwuprzewodowego toru transmisyjnego i różnicowego odbiornika. Przesyłanie różnicowe chroni transmisję danych przed zakłóceniami zewnętrznymi od np. silników i przełączników. RS485 jest często stosowanym interfejsem w sieciach przemysłowych, bo charakteryzuje się dużą szybkością transmisji, nawet na znaczne odległości. W ostatnich latach wzrasta znaczenie połączenia typu przemysłowy Ethernet. W odróżnieniu od innych technologii komunikacyjnych umożliwia funkcjonowanie w jednej sieci wielu różnych protokołów. Wśród najważniejszych można wymienić: EtherCAT, EtherNet/IP, Modbus TCP, Powerlink, Profinet i SERCOS. Przemysłowy Ethernet pozwala na połączenie pomiędzy standardową siecią biurową i siecią przemysłową. Jako medium transmisyjne stosuje się często miedzianą skrętkę ekranowaną. Umożliwia ona na przesyłanie danych klasy Fast Ethernet, czyli z prędkością 100Mb/sekundę z pełnym dupleksem. Maksymalna długość przewodu połączeniowego może wynosić 100m. Coraz ważniejszym medium transmisyjny stają się światłowody, które pozwalają na maksymalny przesyłanie na odległość 14 km. Podłączenie różnych elementów automatyki do sterowania pokazano na rys. 19.



Rys. 20. Cykl programowy sterownika PLC

Kolejne rozkazy sterownika PLC są przetwarzane w sposób cykliczny – po zakończeniu poprzedniego rozkazu jest pobierany następny. W zasadzie nie wykorzystuje się skoków. Gdy zostanie przetworzony ostatni rozkaz z pamięci programu, to następuje zakończenie tzw. cyklu programu, po czym program jest przetwarzany ponownie od początku. Takie działanie jest określane pracą szeregowo-cykliczną, w zamkniętej, nie kończącej się pętli (rys. 20). Sterownik wykonuje zapisany w pamięci program w tzw. cyklu programowym. Na jego początku są wykonywane operacje kontrolne. Należą do nich wszystkie zadania przygotowujące sterownik do kolejnego cyklu niezbędne do jego prawidłowej pracy. Odbywa się wtedy aktualizacja tablicy błędów oraz opóźnienie programowe czasu trwania cyklu (tylko jeśli sterownik jest uruchomiony w trybie stałej długości cyklu). Następnie sterownik przechodzi do wykonywania właściwego programu. Zasadniczo rozkazy są wykonywane jeden po drugim, niezależnie od wyniku poprzedniej operacji. Tylko w wyjątkowych sytuacjach dopuszczalne jest wykonywanie skoku do innych fragmentów programu. Ponieważ czas wykonania pojedynczej instrukcji jest bardzo krótki (od 0,1 do 10 ⌈s), przeto z punktu widzenia procesu wydaje się, że rozkazy wykonywane w rzeczywistości kolejno, są wykonywane równocześnie. Program kończy instrukcja END. Po zakończeniu całej sekwencji następuje faza komunikacji. W jej ramach sterownik nawiązuje komunikację z dołączonymi urządzeniami i wymienia informacje, np. przekazuje dane o stanie we/wy, zmienia parametry itp. Po zakończeniu tej fazy cykl jest

wykonywany od początku. Jednym z ważniejszych parametrów sterownika jest czas cyklu sterownika, składający się z czasu potrzebnego do pełnego obiegu programu złożonego z tysiąca statystycznych instrukcji. Wynosi on od 0,1 do 10 ms.

Odczyt wejść odbywa się jednorazowo na początku cyklu. Stan wejść zostaje zapisany w pamięci sterownika. W następnym kroku sterownik wykonuje program, w trakcie którego pobiera informacje o stanach wejść z pamięci i modyfikuje stan wyjść w pamięci. Po zakończeniu wykonywania programu następuje przepisanie stanów wyjść z pamięci do rzeczywistych modułów wyjściowych. Dzięki takiemu podejściu następuje skrócenie czasu cyklu i brak błędów wynikających ze zmiany sygnałów na wejściach i wyjściach podczas trwania jednego cyklu.

5. Podstawy programowania sterowników PLC

Programowanie w języku drabinkowym

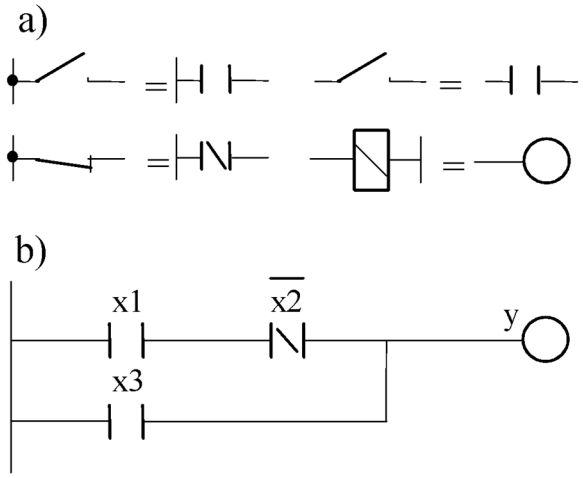
Na wstępie należy zaznaczyć, że każda rodzina sterowników danego producenta ma własną listę rozkazów. Stosuje się następujące języki programowania sterowników binarnych:

- logiki albo schematów drabinkowych (LD),
- bloków funkcyjnych (FBD),
- schematów sekwencyjnych (Sequential Function Chart – SFC)
- listy instrukcji (IL),
- strukturalny, typu język C, Pascal itp.

W Tabeli 2 zamieszczono symbole najczęściej stosowanych w języku drabinkowym elementów. Elementy nr 5 i 6 to wejścia przerzutnika R-S, przy czym elementy (S) oznacza Set – ustaw/zapamiętaj, a element (R) oznacza Reset, czyli kasuj/zeruj. Elementy ze strzałkami tj. nr 7 i 8, ustawiają na wyjściu stan „1” tylko przez jeden cykl obiegu programu odpowiednio, gdy: w poprzednim cyklu było „0” a w obecnym jest „1” albo gdy: w poprzednim cyklu było „1” a w obecnym jest „0”. Element 9 to pamięć – Memory (M).

	Tabela 2
–| |–	Styk normalnie otwarty
–| / |–	Normalnie zamknięty)
–( )–	Cewka
–(/)–	Cewka negująca
–(S)–	Cewka ustawiająca
–(R)–	Cewka kasująca
–(□)–	Cewka– zbocze narastające
–(□)–	Cewka– zbocze opadające
–(M)–	Cewka pamięci

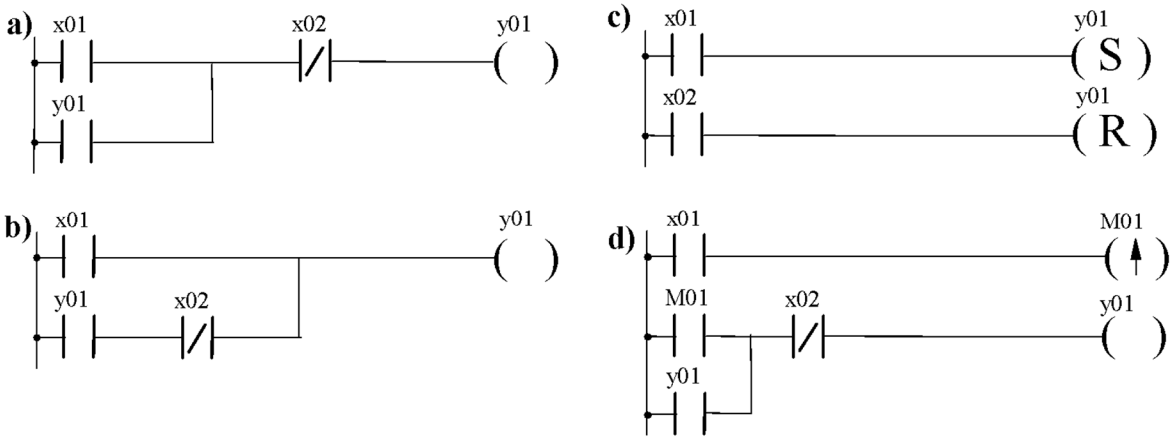
Język logiki drabinkowej jest językiem graficznym, który stworzono pierwotnie, aby zastąpić układy przekaźnikowe. Z pewnym uproszczeniem można uznać, że logika drabinkowa polega na narysowaniu schematu sterowania przekaźnikowego na ekranie programatora. Wynika to z faktu, iż sterowniki zastąpiły używane wcześniej panele przekaźnikowe. Język ten jest łatwo przyswajany i zrozumiały dla personelu obsługującego. Opiera się on na standaryzowanych symbolach graficznych, odpowiadających elementom stykowym, czyli różnym zestykom i cewkom przekaźników. Przy programowaniu, schemat stykowo-przekaźnikowy należy obrócić o 90º. Możliwe jest dodatkowo używanie funkcji: arytmetycznych i logicznych oraz porównań i relacji, a także bloków funkcyjnych, takich jak np.: przerzutniki, czasomierze, liczniki, regulatory PID i inne. Język schematów drabinkowych jest językiem powszechnie używanym i uwzględnionym w normach IEC1.



Rys. 21. Symbole graficzne stosowane w języku LD (a) oraz przykładowy program (b)

Na rysunku 21a pokazano zestyk i cewkę przekaźnika oraz odpowiadające im symbole graficzne stosowane w języku LD. Są one rysowane poziomo – tak jak stopnie drabiny, stąd też potoczna nazwa języka LD jako schemat drabinkowy. Podstawowym elementem obwodu jest styk, który może być zwierny (normalnie otwarty) lub rozwierny (normalnie zamknięty). Przykładowy program napisany w języku LD pokazano na rys. 21b. Poszczególne szczeble schematu (linie poziome) są nazywane obwodami i składają się z połączonych ze sobą pojedynczych elementów wejściowych (zestyków). Na końcu każdej linii musi występować element wykonawczy (wyjście binarne), np. cewka, przerzutnik, licznik, pamięć itp. Obwód jest ograniczony z lewej strony szyną zasilania tj. 24VDC. Zasadą działania języka LD jest przepływ wirtualnego prądu od linii zasilania przez poszczególne modele elementów do

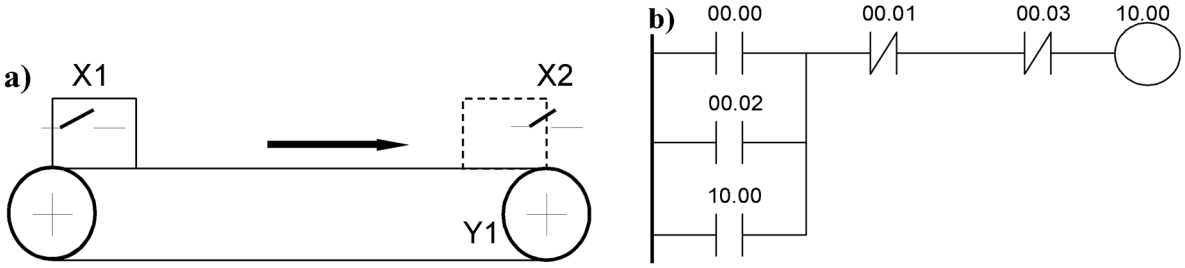
elementy wyjściowego. Element wyjściowy zostaje ustawiony w odpowiedni stan (1 albo 0) w zależności od tego, czy ten wirtualny prąd odpowiednio dopływa do element wykonawczego czy też nie dopływa. Program jest wykonywany krok po kroku (obwód po obwodzie) aż do ostatniej linii albo do instrukcji END. Wyjątkiem od tej reguły są to tzw. skoki, które mogą być warunkowe, postaci, jeśli warunek, to skok do etykieta lub bezwarunkowe, postaci skok do etykieta. Jednak skoki w programach występują bardzo rzadko. Przy programowaniu algorytmów sterowania konieczne jest wykonywanie różnorodnych operacji uzależnionych od czasu, typu opóźnienia. Mogą one być programowane z wykorzystaniem elementów czasowych. Ważne są również moduły licznikowe, które pozwalają na zliczanie różnorodnych, występujących w procesie stanów.

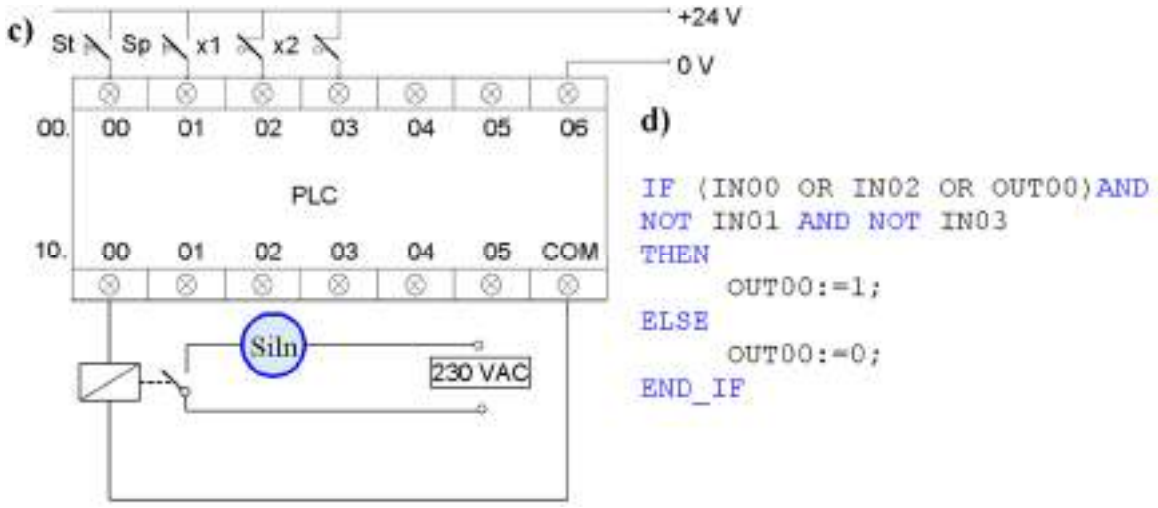


Rys. 22. Układy pamięci: a) z przewagą kasowania, b) z przewaga ustawiania, c) z wykorzystanie elementów S i R, d) aktywowanej zboczem narastającym

Na rysunku 22a zamieszczono program realizujący funkcję pamięci z podtrzymaniem, opisaną poniższym równaniem:

$$y01 = (x01 + y01) \cdot \overline{x02}$$





Rys. 23. Sterowanie taśmociągiem: a) rysunek taśmociągu, b) program w języku LD, c) schemat podłączenia do sterownika elementów, d) program w języku strukturalnym

Na rysunku 23 przedstawiono układ do sterowania, do uruchomiania silnika taśmociągu, po pojawieniu się przedmiotu na jego początku, co jest sygnalizowane czujnikiem x1 (obecność przedmiotu = zwarcie, czyli „1") lub wciśnięcie startu przełącznikiem St (zwarcie). Taśmociąg powinien zatrzymać się po osiągnięciu pozycji końcowej przez przemieszczany przedmiot, sygnalizowane czujnikiem x2 (obecność przedmiotu = zwarcie) lub naciśnięciem stopu awaryjnego przyciskiem Sp. Na rysunku 23c pokazano schemat połączeń do sterownika PLC, a na rys. 23d program sterowania.

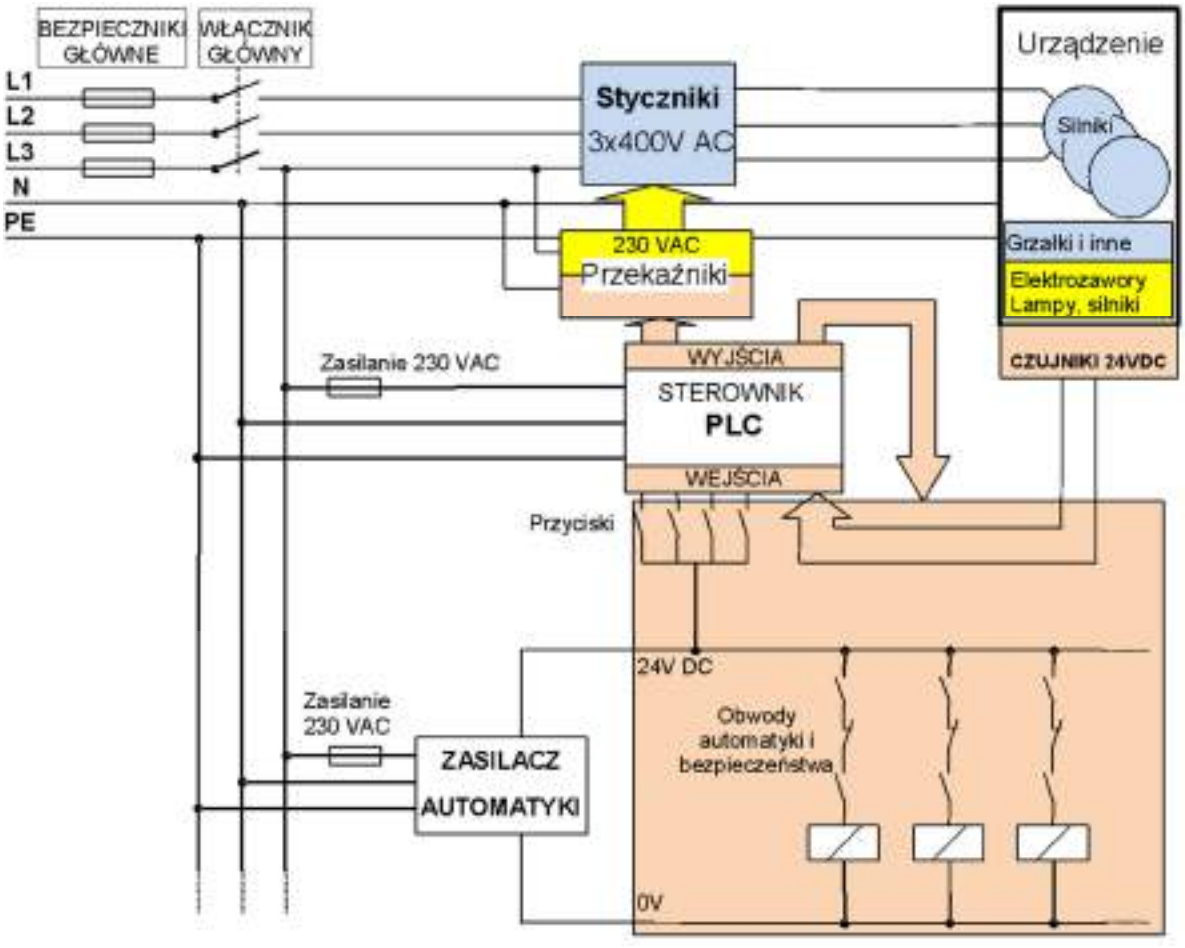
Zasilanie i połączenia urządzeń zautomatyzowanych

Sterowniki przemysłowe, różne czujniki oraz elementy wykonawcze małej mocy są zwykle zasilane napięciem 24 VDC. Takie napięcie jest zarówno bezpieczne dla człowieka, jak i odpowiednio wysokie dla zapewnienia ochrony układów automatyki przed zakłóceniami. Pozwala na włączanie i wyłączanie elementów o mocy do ok. 100 W, takich jak elektrozawory, przekaźniki, małe silniki, diody sygnalizacyjne itp. Przekaźniki są najczęściej wykorzystywane do włączania urządzeń o mocach do 2-3 kW, zasilanych napięciem jednofazowym 230 VAC. Do włączania i wyłączania elementów o większej mocy stosowane są zwykle trójfazowe styczniki.

Zabezpieczenie silników zarówno 1- jak i 3-fazowych jest niezbędne we wszystkich układach w automatyce przemysłowej. Systemy ochronne, mają za zadanie wykrywanie i reagowanie na potencjalne problemy, zanim doprowadzą one do uszkodzenia układu.

Zabezpieczenia monitorują funkcjonowanie silników 3-fazowych, śledząc takie parametry jak natężenia prądu, napięcia i temperaturę. Zastosowanie zabezpieczeń dla silników 3-fazowych minimalizują ryzyko przeciążeń i skoków napięcia, które mogą grozić uszkodzeniem silnika. Ponadto, zabezpieczenia te zwiększają bezpieczeństwo procesów przemysłowych, chroniąc personel i maszyny przed potencjalnymi niebezpieczeństwami. Ich istotną funkcją jest także poprawa wydajności pracy silników, co przekłada się na zwiększenie wydajności całego systemu przemysłowego. Niezabezpieczony silnik narażony jest na różne uszkodzenia powodowane przeciążeniami, zwarciami czy brakiem fazy. Stosowanymi urządzeniami do zabezpieczenia silnika trójfazowego są:

- wyłącznik różnicowo-prądowy – jest stosowany, gdy silnik zasilany jest bezpośrednio prądem z sieci tzn. nie ma falownika; zapewnia ochronę przeciwporażeniową;
- detektor zaniku fazy – zapobiega przeciążeniom spowodowanym przez zanik jednej z faz podczas pracy urządzenia;



Rys. 24. Schemat układ zasilania elementów automatyki

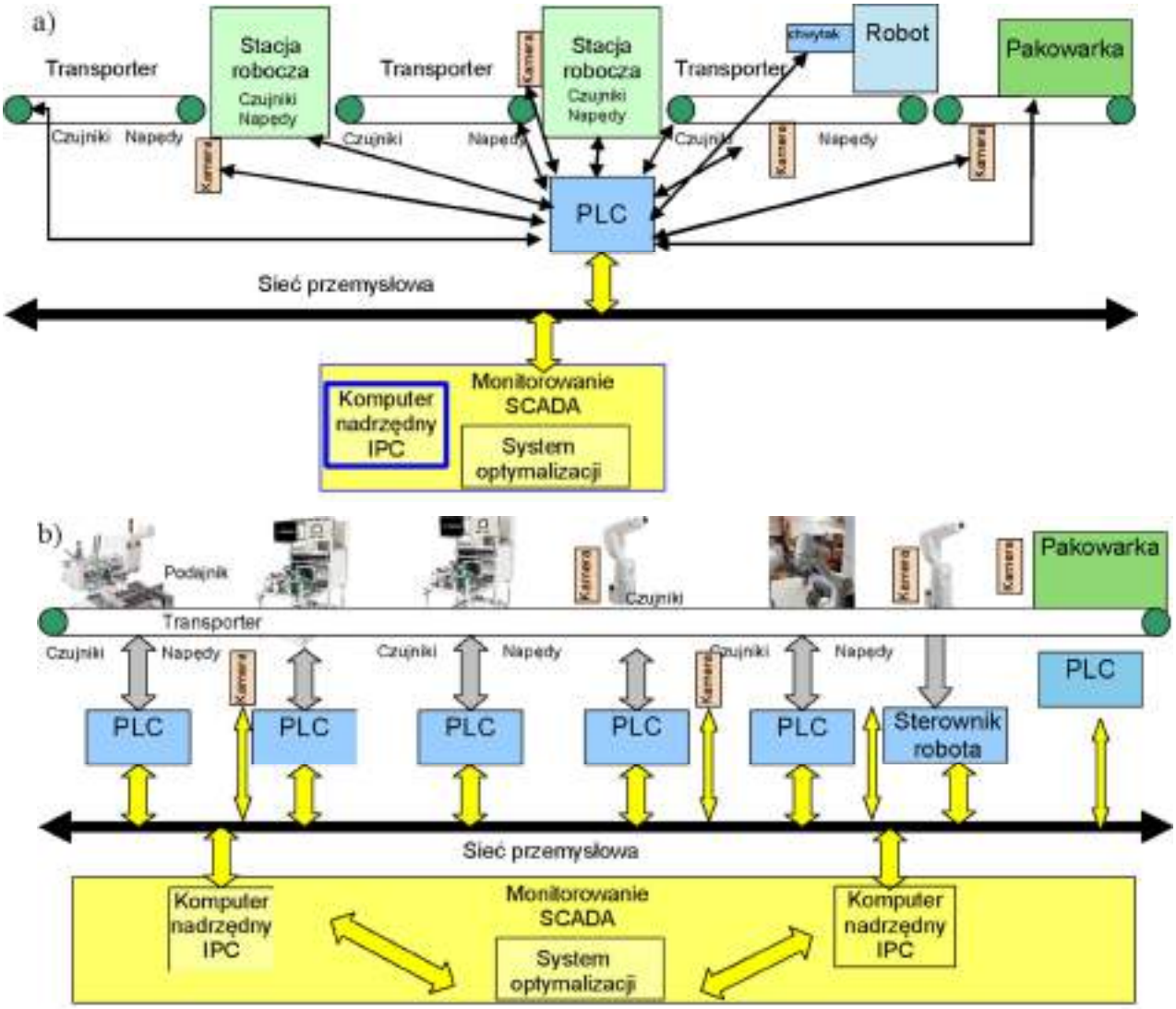
- wyłączniki silnikowe – wyłączają i załączają silniki; wyposażone są w wyzwalacz termiczny i elektromagnetyczny. Wyłączają silnik w ciągu kilku milisekund, gdy dojdzie do przekroczenia jego prądu znamionowego;
- czujnik zaniku fazy – łączy się ze stycznikiem albo z wyłącznikiem silnikowym; zabezpiecza silnik przed uszkodzeniem, spowodowanym w przypadku działania tylko dwóch faz. Ma za zadanie rozłączać obwód, gdy braknie jednej z faz;
- czujnik asymetrii napięć zasilania – wyłącza zasilanie, gdy występuje różnica w ich wartościach napięć lub pojawi się niewłaściwe przesunięcie kątów.

Zastosowanie omówionych urządzeń pomaga skutecznie chronić silnik przed spaleniem czy innym uszkodzeniem na wypadek zaistnienia nieprawidłowości w sieci zasilającej lub instalacji elektrycznej. Oczywiście, aby sprawnie działały i należycie spełniały swoją funkcję, należy je dobrać do mocy silnika i parametrów układu. Schemat układu zasilania urządzenia zautomatyzowanego pokazano na rys. 24.

6. Podstawy projektowania linii zautomatyzowanych

Budowa zautomatyzowanej linii produkcyjnej

Zautomatyzowana linia produkcyjna to zaawansowany system, w którym zastąpiono pracę ludzi maszynami i urządzeniami. Proces produkcyjny jest podzielony na etapy, wykonywane przez automatyczne urządzenia, maszyny i roboty, które realizują takie zadania jak podawanie części i surowca, obróbka-przetwarzanie, montaż i kontrola jakości, a niekiedy obejmuje też pakowanie. Poprawnie zaprojektowania linia, zapewnia płynność produkcji oraz efektywność i powtarzalność. Linie produkcyjne integrują maszyny, roboty, systemy komputerowe i oprogramowanie kontrolne, aby automatyzować różne etapy procesu produkcyjnego.



Rys. 25. Schematy zautomatyzowanej linii produkcyjnej: a) z wieloma transporterami, b) z transporterem centralnym (produkcja ciągła)

Dzięki automatyzacji zwiększeniu ulega zarówno wydajność, jak i precyzja produkcji, a zautomatyzowanych liniach produkcyjnych (rys. 25) stosowany jest zestaw urządzeń, które zapewniają proces wytwarzania, nie wymagający uczestnictwa człowieka. W zależności od wykonywanej funkcji na linii produkcyjnej można wyróżnić kilka grup urządzeń, tj.:

a. Elementy wykonawcze:

- zautomatyzowane systemy przenośnikowe np. taśmowe, rolkowe, łańcuchowe, transportery, podajniki itp.,
- stacje robocze:
  - odwijarki blachy albo drutu, prasy, wtryskarki, zgrzewarki, spawarki (procesy formowania i łączenia materiałów)
  - obrabiarki, w tym CNC tj.: przecinarki, wiertarki, frezarki, tokarki, szlifierki,

- roboty przemysłowe (spawanie, montaż, przenoszenie, paletyzacja, malowanie)
- podnośniki, chwytaki, dociskarki – precyzyjne podnoszenie i ustawianie detali.
- Napędy pneumatyczne, hydrauliczne i elektryczne (pozycjonowanie).

b. Elementy sterujące – tworzące tzw. „mózg" linii:

- PLC – sterowniki przemysłowe do sterowania logiką pracy całej linii, do których podłączone są elementy pomiarowe tj. czujniki, przyciski, termopary, enkodery itp. oraz elementy wykonawcze tj. grzałki, elektrozawory, napędy itp.
- IPC – komputery przemysłowe do nadzorowania, wizualizacji i obsługi złożonych algorytmów, systemów wizyjnych, systemów SI.
- HMI – panele operatorskie, tworzące interfejs człowiek-maszyna, przeznaczony do kontroli, wizualizacji i nastawiania parametrów procesu, alarmowania itp.
- SCADA, MES, ERP – systemy informatyczne do nadzorowania produkcji, raportowania itp.

c. Elementy pomiarowe i kontrolne, tworzące tzw. „zmysły" linii

- czujniki 0/1 np. optyczne, laserowe, indukcyjne, pojemnościowe, ultradźwiękowe – do wykrywania obecności,
- termopary, termorezystory, wagi, przepływomierze, momentomierze, tensometry – stosowane do pomiaru: temperatury, masy, objętości, siły itp.
- potencjometry, enkodery – do precyzyjnego pomiaru pozycji
- kamery 2D/3D, RFID, systemy wizyjne – do kontroli obecności, jakości, odczytu kodów, pozycjonowani.

d. Elementy napędowe

- silniki elektryczne – AC, DC, krokowe.
- napędy i serwonapędy z silnikami indukcyjnymi, BLDC, PMSM Falowniki serwa— regulacja prędkości i momentu.
- siłowniki pneumatyczne i hydrauliczne – do szybkiego przesunięcia i generowania dużych sił (do podnoszenia, docisku, obróbki plastycznej).
- zawory, rozdzielacze, pompy — sterowanie mediami roboczymi.

e. Elementy komunikacyjne i integracyjne

- sieci przemysłowe – EtherCAT, Modbus, Profinet, CANopen itp.
- chmura przemysłowa / IoT — analiza danych, predykcja awarii, optymalizacja,
- Systemy do śledzenia traceability – zapis parametrów w trakcie produkcji

f. Elementy pomocnicze

- Stacje pakujące i etykietujące — automatyczne przygotowanie produktu końcowego
- Bufory i magazyny zautomatyzowane do składowania i balansowania linii
- Podajniki np. wibracyjne, sortowniki do orientowania i podawania małych detali.

g. Elementy bezpieczeństwa

- Osłony, bariery, zamki bezpieczeństwa
- Kurtyny świetlne do zatrzymania produkcji po wejściu człowieka
- Systemy i sterowniki bezpieczeństwa.

Projektowanie zautomatyzowanej linii produkcyjnej

Przed przystąpieniem do projektowania automatyzacji linii produkcyjnej, trzeba rozpoznać jej budowę, działanie itd. Projektowanie należy realizować w kilku krokach (rys. 26):

1. Rozpoznać ogólną budowę i działanie automatyzowanych poszczególnych urządzeń i całej linii produkcyjnej, w szczególności zadań automatyzacji:
   – ustalić zasady (funkcje) według których mają działać poszczególne urządzenia zautomatyzowane
   – zaproponować tabele prawdy, grafy reguły włączania i wyłączania poszczególnych urządzeń
   – określić reguły sterowania cała liną produkcyjną.
2. Rozpoznać ogólnie zadania urządzeń/linii produkcyjnej i ich parametry (wymagane) tj.:
   – wydajność, szybkość (np. ruchu)
   – dokładność wykonania
   – wymagania dotyczące niezawodności
   – określić co (jakie wielkości fizyczne) trzeba będzie mierzyć i z jaką dokładnością i precyzją
   – zdefiniować jakie trzeba będzie zastosować silniki napędy w zakresie: wymaganych mocy średniej (znamionowej) i maksymalnej (chwilowej), momentów i sił, prędkości, zakresu ruchu, dokładności regulacji siły, prędkości i przyspieszenia
   – określić wstępne wymagania dotyczące zastosowania sterowników w zakresie:
   • jakimi elementami i urządzeniami trzeba sterować
   • które zadania mają charakter sterowania ciągłego, a które binarnego
   • jakie algorytmy będzie można zastosować.

3. Dobrać metody pomiarów oraz czujniki i urządzenia pomiarowe
   - określić, gdzie tj. w których miejscach zastosować czujniki binarne
   - określić wymagania stawiane tym czujnikom (co wykrywać, z jakiej odległości, jaka będzie maksymalna częstotliwość, itp.)
   - dobrać czujniki obecności (jest/nie ma - binarne: 0/1)
   - określić jakie wielkości fizyczne mierzyć (przemieszczenie, prędkość, siła, temperatura …) i z jaką dokładnością oraz z jakim czasem pomiaru
   - dobrać jakie urządzenia pomiarowe zastosować, np. enkodery, liniały pomiarowe, wagi tensometryczne, czujniki laserowe, profilometry, termopary itp.
   - określić układy połączeń wybranych czujników i urządzeń ze sterownikami.
4. Dobrać zespoły napędowe
   - jakie, ile i gdzie zastosować silniki, napędy itp.
   - określić ich wymagane parametry, a w szczególności czy powinna być stosowana regulacja np. prędkości i pozycjonowania
   - wybrać rodzaj napędów: elektryczne, hydrauliczne, pneumatyczne
   - ustalić wymagane: moc, prędkości, siły, momenty, dokładności
   - dobrać napędy i ich układy załączania i regulacji
   - wybrać sposób komunikacji napędów ze sterwnikami
5. Dobrać sterowniki PLC albo PAC
   - określić liczbę wejść i wyjść binarnych oraz analogowych
   - określić inne wymagane moduły np. regulatory, przetworniki itp.
   - określić wymagania dotyczące sterowników w zakresie szybkości działania
   - wybrać sieć komunikacyjną
   - dobrać sterowniki napędów, a w szczególności sposób ich komunikacji z PLC
6. Rozpoznać wymagania dotyczące: minimalizacji zużycia energii, BHP, diagnostyki itp.

Rys. 26. Widok elementów automatyki stosowanych na zautomatyzowanej linii produkcyjnej

Sterowniki PLC, kontrolery PAC oraz panele operatorskie i komputery panelowe to istotne elementy nowoczesnej automatyki przemysłowej. Ich rola w automatyzacji, w tym w Przemyśle 4.0 stale rośnie, a użytkownicy oczekują coraz większej wydajności, elastyczności i zaawansowanych funkcji komunikacyjnych. Systemy z PLC i PAC pozostają filarem automatyki dzięki niezawodności i łatwej integracji. Współpracują one z panelami HMI (Human-Machine-Interface), które odgrywają ważną rolę w wizualizacji i sterowaniu procesami przez operatorów. Rozwiązania te znajdują zastosowanie we wszystkich gałęziach przemysłu tj. w produkcji, energetyce, transporcie i logistyce. Ich rozwój wprowadza cyfryzację i możliwości dogłębnej analizy danych w czasie rzeczywistym. Ważnym trendem jest wykorzystanie algorytmów bazujących na sztucznej inteligencji. Istotnym problemem, ograniczającym powszechny postęp w zakresie cyfryzacji i automatyzacji, w tym pełnej koncepcji Przemysłu 4.0, jest konieczność ochrony sterowników programowalnych przed nieautoryzowanym dostępem i złośliwym oprogramowaniem, która jest konieczna w reakcji na rosnące zagrożenie cyberatakami. Stosowane obecnie sterowniki przemysłowe PLC można podzielić na kilka grup: nanosterowniki (do 32 we/wy), mikrosterowniki (do 128 we/wy), średniej wielkości (do 1024 we/wy) oraz duże (> 1024 we/wy do nawet 10 000). Oprócz tego podział sterowników PLC obejmuje:

- sterowniki kompaktowe/modułowe

- sterowniki do systemów rozproszonych/do pracy wieloprocesorowej
- wersje z redundancją/typu fail-safe
- wersje z web serwerami/z OPC.

Najważniejszym kryterium doboru sterownika PLC jest to, aby można było podłączyć do niego wszystkie elementy, konieczne do zrealizowania zadania automatyzacji. W pierwszym rzędzie chodzi o odpowiednie liczby wejść i wyjść, tak aby każdy element bądź urządzenie mogły być do PLC podłączone i obsłużone programowo. Ponadto dobrany sterownik powinien mieć możliwość rozszerzenia swoich możliwości, w tym zwiększenia liczby wejść/wyjść, aby umożliwić ewentualną rozbudowę.

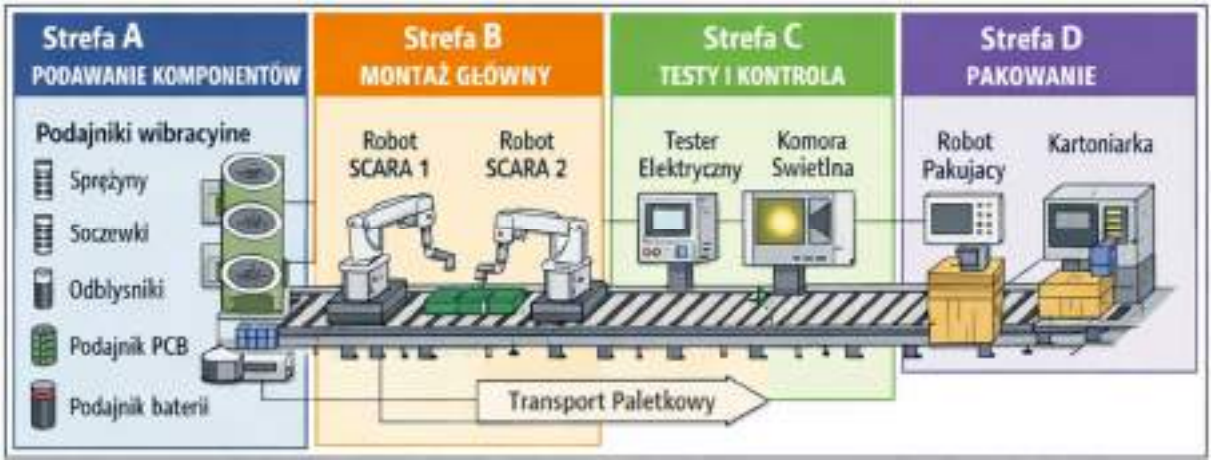
Poniżej przedstawiono, przykładowe koncepcje zautomatyzowania kilku linii przemysłowych. W ich przygotowaniu wykorzystano sztuczną inteligencję w postaci programów ChatGPT, Copilot i Gemini.

Przykład 1: Zautomatyzowana linia montażu lampki bateryjnej.

1. Założenia
   - Elementy: obudowa dolna, obudowa górna, płytka PCB z diodą LED, płytka montażowa, soczewka, przycisk, sprężyna, bateria.
   - Wymagania: 1200 szt./zmianę → takt ok. 24 sekundy.
   - Kontrola jakości: test elektryczny, test światła, weryfikacja obecności elementów.

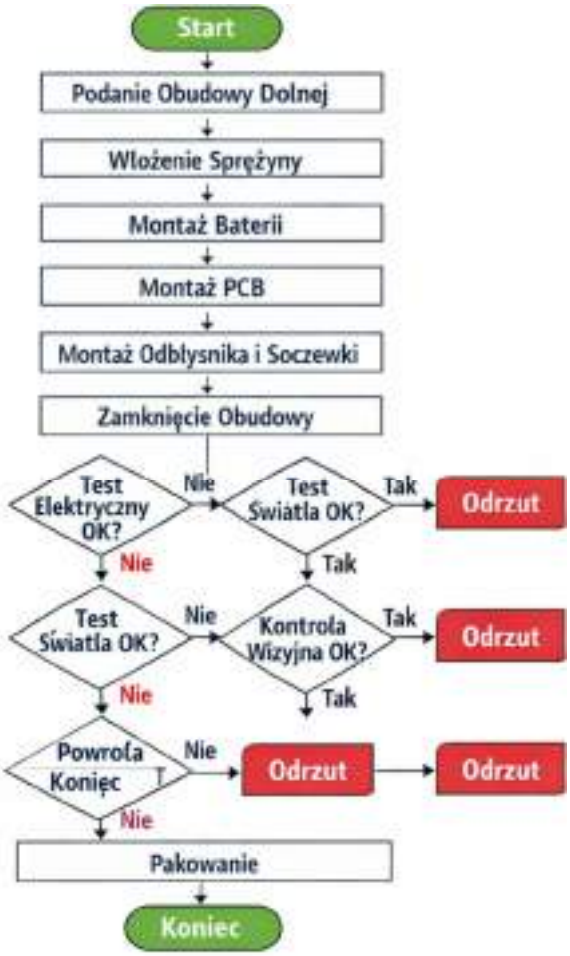
2. Przepływ procesu – pełna sekwencja operacji
   1. Podanie obudowy dolnej – podajnik wibracyjny + pozycjonowanie.
   2. Włożenie sprężyny – robot SCARA pobiera sprężynę i wkłada ją do obudowy.
   3. Montaż baterii – podajnik rolkowy podaje baterię, manipulator umieszcza ją w obudowie.
   4. Podanie PCB – podajnik tackowy, pobranie przez robota.
   5. Włożenie PCB i dociśnięcie – robot umieszcza płytkę i dociska do zatrzasków.
   6. Montaż płytki i soczewki – dwa podajniki wibracyjne, robot montuje elementy.
   7. Podanie obudowy górnej – podajnik wibracyjny.
   8. Zamknięcie obudowy – docisk pneumatyczny z kontrolą siły i drogi.
   9. Test elektryczny – automatyczne przyłożenie styków, pomiar prądu i napięcia.
   10. Test światła – fotodioda mierzy natężenie światła.
   11. Kontrola wizyjna – kamera sprawdza kompletność i poprawność montażu.

12. Odrzut wadliwych sztuk – bramka pneumatyczna.
13. Pakowanie jednostkowe – robot odkłada produkt do tacki lub kartonika.
14. Pakowanie zbiorcze – kartonowanie 20 sztuk.



Rys. 27. Zautomatyzowana linia do montażu lampek (ChatGpt)

3. Rozmieszczenie stacji/gniazd na linii (layout – rys. 27)



Rys. 28. Algorytm procesu montażu

- Strefa A – podawanie komponentów przez podajniki wibracyjne: sprężyna, soczewka, odbłyśnik, obudowy. Podajnik tackowy: PCB. Podajnik rolkowy: baterie.
- Strefa B – montaż główny. Dwa roboty SCARA + prasa pneumatyczna.
- Strefa C – testy i kontrola. Tester, kamera wizyjna.
- Strefa D – pakowanie. Robot pick&place + kartoniarka.

4. Dobór urządzeń
   - Roboty SCARA.
   - Podajniki wibracyjne.
   - Prasa pneumatyczna – z czujnikiem siły i drogi, aby wykrywać niepełne zatrzaśnięcia.
   - System wizyjny – kamera 5 MP, oświetlacz pierścieniowy, detekcja braków.
   - Tester elektryczny – pomiar prądu diody.
   - Kamera – pomiar natężenia światła.
   - PLC + HMI – sterowanie całą linią.
5. Algorytm sterowania (rys. 28)
   1. PLC odbiera sygnał „obudowa dolna gotowa”.
   2. Robot 1 pobiera sprężynę i montuje.
   3. Robot 1 pobiera baterię i montuje.
   4. Robot 2 pobiera PCB i montuje.
   5. Robot 2 montuje soczewkę.
   6. Prasa zamyka obudowę – jeśli siła/droga poza tolerancją → odrzut.
   7. Tester elektryczny uruchamia pomiar – jeśli prąd nieprawidłowy → odrzut.
   8. Komora świetlna mierzy natężenie – jeśli poniżej progu → odrzut.
   9. Kamera wizyjna sprawdza kompletność – jeśli błąd → odrzut.
   10. Robot pakujący odkłada produkt.

Przykład 2: Zautomatyzowana linia do produkcji przednich lamp samochodowych LED

W ostatnich latach do oświetlenia pojazdów w zakresie lamp przednich stosowane są lampy typu LED. Ta technologia niemal całkowicie zastąpiła tradycyjne źródła światła w reflektorach w postaci żarówek, lamp halogenowych oraz ksenonowych. Technologia LED jest stosowana głównie ze względu na jej większą trwałość i niższe zużycie energii. Daje ona także duże możliwości uzyskania zaawansowanych układów optycznych. Najważniejsze zalety stosowanej w motoryzacji technologii lamp typu LED są następujące:

- wysoka sprawność energetyczna, co oznacza małe zużycie energii
- długa żywotność wynosząca ok. 30 000 godz.
- duża odporność na wstrząsy i przeciążenia
- możliwość tworzenia adaptacyjnych systemów oświetlenia
- mniejsze rozmiary elementów świetlnych.

Sama produkcja lamp samochodowych ma zwykle charakter masowy, bowiem producenci samochodów potrzebują rocznie ponad 100 000 do 1 000 000 lamp jednego rodzaju. W związku z tym wymagana jest zarówno duża zdolność produkcyjna lamp jak i wysoka powtarzalność, a także duża dokładność montażu oraz zapewnienia kontroli jakości. Dlatego w zakładach produkujących lampy LED stosuje się bardzo zautomatyzowane linie produkcyjne, w tym wyposażone w stacje zrobotyzowane. Zakłada się następujące parametry: produkcja dzienna: 200000/250 = 800 szt./dzień, produkcja na zmianę: 800/2 = 400 szt./zmianę, produkcja na godzinę: 400/8 = 50 szt./h, czas cyklu produkcyjnego 3600 s/50 = 72 s ≈ 70 s.

Rys. 29. Elementy składowe lampy LED

Zakłada się, że projektowana zautomatyzowana linia produkcyjna będzie zdolna do wytwarzania 200 000 sztuk rocznie przednich lamp samochodowych LED. Poniżej przedstawiono jej cechy charakterystyczne. Przyjęto, że produkowana będzie przednia lampa LED, która składa się z kilku następujących podstawowych elementów konstrukcyjnych:

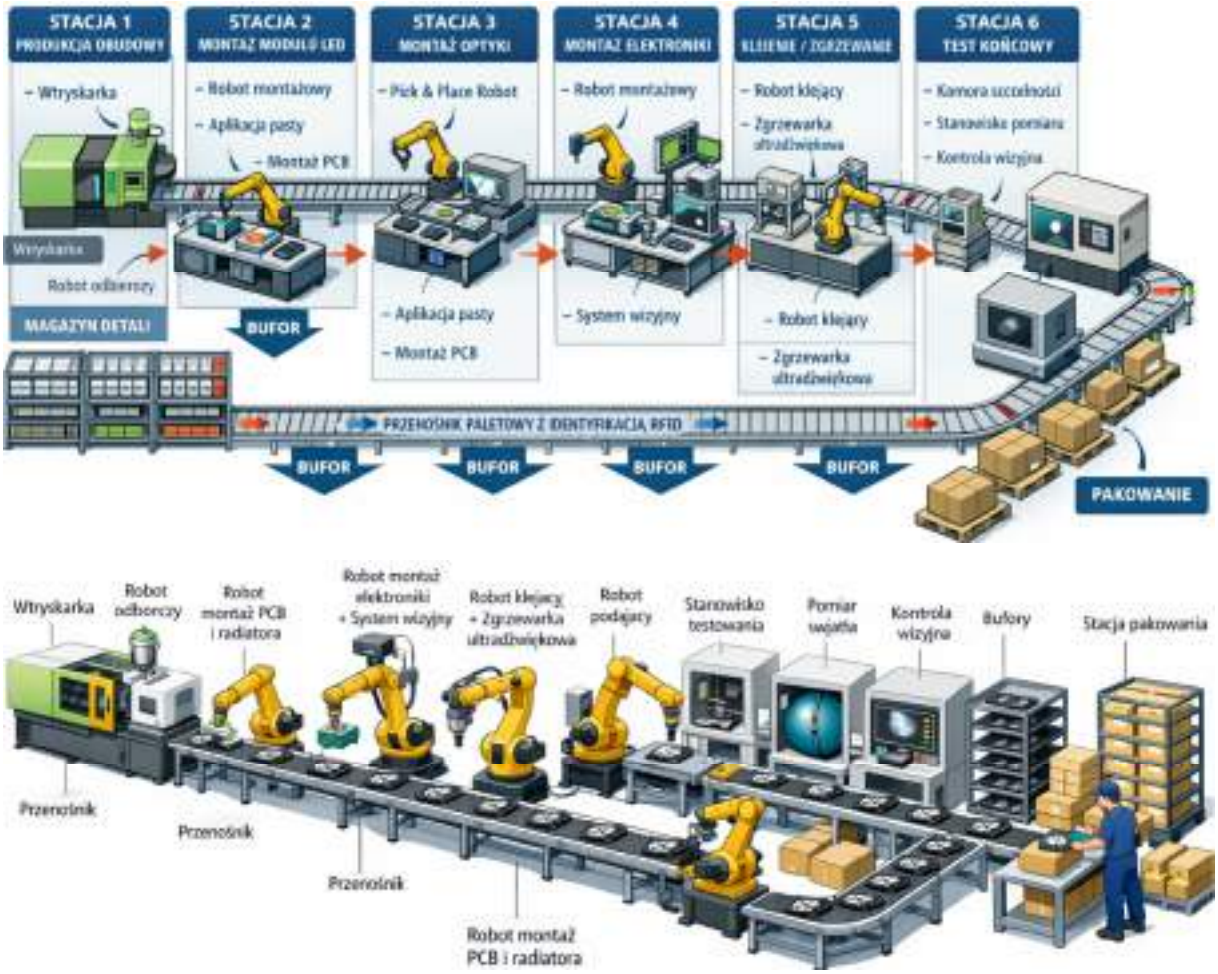
- układu optycznego składającego się z diody LED z radiatorem
- soczewki/projektora,
- odbłyśnika/reflektora,
- sterownika elektronicznego i przewodem z wtyczką zasilania,
- obudowy lampy,
- uszczelnienia.

Proces produkcji składa się z kilku etapów technologicznych. Pierwszym elementem jest produkcja obudowy lampy, która wytwarzana jest zwykle metodą wtrysku tworzyw sztucznych, z materiałów takich jak: ABS, PP, PC-ABS. Proces montażu całej lampy LED będzie obejmował (rys. 30):

1. Produkcja obudowy lampy
   Operacje: uplastycznienie granulatu, wtrysk do formy, chłodzenie, wyjęcie detalu.
   Zastosowano: wtryskarkę, robota do odbioru obudów oraz przenośnika
   Przyjęto, że cykl wtrysku będzie wynosił około: 60–80 s.
2. Zamontowanie modułu LED i radiatora
   Będzie on służył do odprowadzenia ciepła. Proces będzie wykonywany przez robota przemysłowego, w następujących etapach: podanie radiatora, aplikacja pasty termoprzewodzącej, montaż płytki PCB, przykręcenie śrub.
   Zastosowane będą urządzenia: robot przemysłowy, dozownik pasty, wkrętarka automatyczna
3. Montowanie elementów optycznych lampy tj.: soczewki i reflektora.
   Ta operacja będzie wymagała dość wysokiej dokładności ustawienia obu elementów.
   Zastosowane będą: robot pick and place, system wizyjny
4. Instalacja elektronicznego modułu sterującego zasilaniem diody LED,
   Będzie on zmieniał parametry elektryczne, a tym samym rodzaj oświetlenia drogi.
   Będzie wykonywany w następujących krokach: montaż sterownika, podłączenie przewodów, wykonanie testu elektrycznego – sprawdzenie działania.
   Zastosowane będą: robot montażowy i stanowisko testowe
5. Proces zamknięcia lampy, tzn. połączenia jej klosza z obudową.
   Będzie to wykonywane za pomocą: klejenia albo zgrzewania ultradźwiękowego.
   Urządzenia: dozownik kleju i zgrzewarka ultradźwiękowa
6. Test końcowy. Po zakończeniu montażu każda lampa przejdzie, który będzie obejmował: test szczelności, test elektryczny, pomiar natężenia światła, kontrola wizualna.

Zaplanowano zastosowanie buforów, które umożliwią stabilizację pracy linii produkcyjnej. Na linii zaplanowano zastosowanie robotów przemysłowych, a ich łączna liczba będzie wynosić od 6 do 8. Ich funkcje to: transport elementów, montaż komponentów, aplikacja kleju, obsługa stanowisk testowych. Kontrola jakości realizowana będzie na kilku etapach produkcji. Będzie ona obejmowała: kontrolę wizyjną, elektryczną, pomiar parametrów

świetlnych oraz test szczelności. Dzięki temu możliwe będzie wykrycie wad na wczesnym etapie produkcji.



Rys. 30. Dwie różne wizualizacje linii do produkcji lamp samochodowych (wg ChatGPT)

W stacjach zastosowane będą następujące maszyny i urządzenia (rys. 30):

Stacja 1 – produkcja obudowy. Maszyny: wtryskarka, robot odbiorczy, przenośnik

Stacja 2 – montaż radiatora i modułu LED. Urządzenia: robot przemysłowy, dozownik pasty, wkrętarka automatyczna

Stacja 3 – montaż optyki. Urządzenia: robot pick and place, system wizyjny

Stacja 4 – montaż elektroniki. Urządzenia: robot montażowy, stanowisko testowe

Stacja 5 – zamknięcie lampy. Urządzenia: dozownik kleju, zgrzewarka ultradźwiękowa

Stacja 6 – test końcowy. Urządzenia: komora szczelności, pomiaru światła, system wizyjny.

6. Kontrola jakości: wizyjna, elektryczna, pomiar parametrów świetlnych, szczelności

7. System transportowy. Urządzenia: przenośnik paletowy, bufory, system i RFID.

W tabeli 3 przedstawiono zestawienie podstawowych elementów, które zostaną zastosowane.

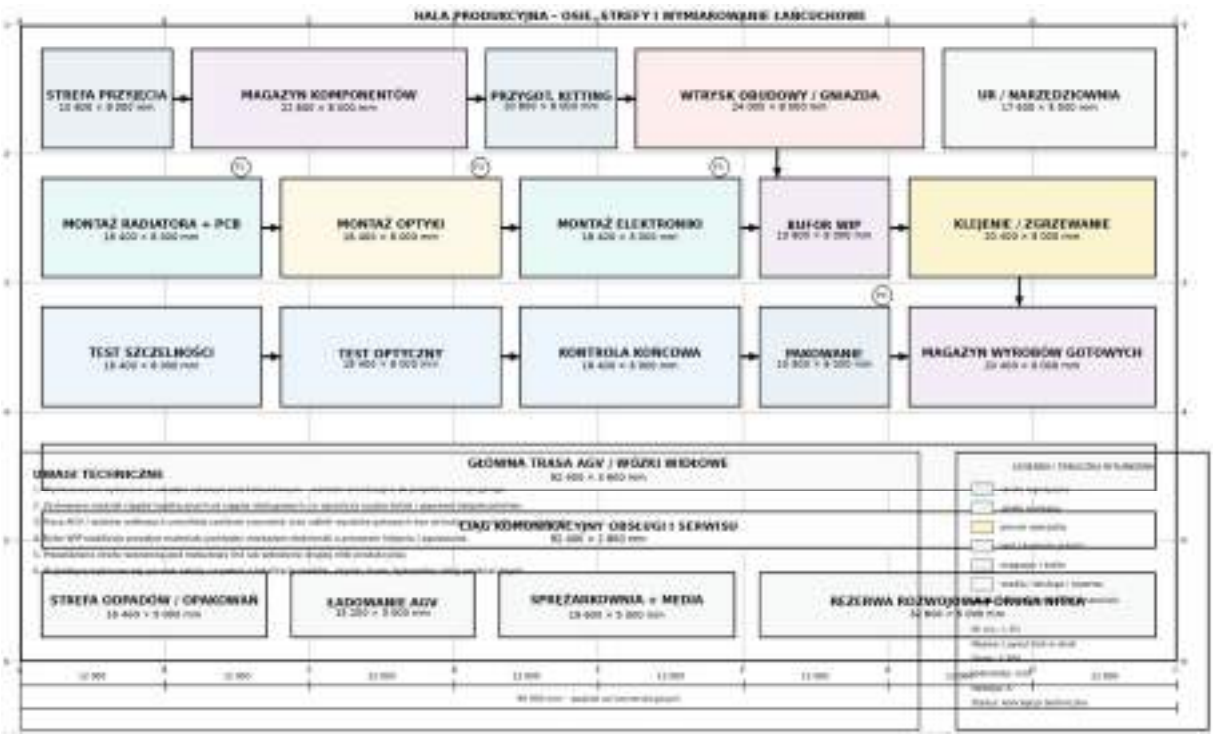
Tabela 3. Zestawienie elementów automatyki

Lp.	Elem. automatyki model / typ	Producent	Ilość	Zastosowanie
1	Sterownik PLC /S7-1515-2 PN	Siemens	1	Centralne sterowanie linią produkcyjną
2	Moduł we cyfr/ SM521 DI 32x24V	Siemens	2	Odczyt sygnałów z czujników
3	Moduł wyj cyfr/ SM522 DO 32x24V	Siemens	2	Sterowanie elementami wykonawczymi
4	Moduł we analog / SM531 AI	Siemens	1	Odczyt sygnałów analogowych
5	Panel operat. HMI / TP1200 Comfort	Siemens	2	Sterowanie i wizualizacja pracy linii
6	Zasilacz przem / SITOP PSU 24V	Siemens	2	Zasilanie układów sterowania
7	Robot przemysłowy / IRB 1200	ABB	4	Montaż elementów lampy
8	Robot pick and place / KR10 R1100	KUKA	2	Montaż optyki
9	Sterownik robota / IRC5	ABB	4	Sterowanie robotem
10	System wizyjny / In-Sight 7000	Cognex	3	Kontrola jakości montażu
11	Kamera przemysłowa / CV-X	Keyence	2	Inspekcja wizualna
12	Czujnik indukcyjny / IGK3008	IFM	20	Wykrywanie elementów metalowych
13	Czujnik optyczny / WL12-3	SICK	25	Detekcja obecności detalu
14	Czujnik wizyjny / IV2	Keyence	5	Kontrola poprawności montażu
15	Falownik / SINAMICS G120	Siemens	10	Sterowanie prędkością przenośników
16	Silnik elektryczny/ DRN80M4	SEW	10	Napęd przenośników
17	Serwonapęd/ SINAMICS S210	Siemens	4	Precyzyjne sterowanie manipulatorów
18	Kurtyna świetlna / deTec4	SICK	6	Ochrona operatorów
19	Sterownik bezpieczeństwa/ S7-1500F	Siemens	1	System bezpieczeństwa
20	Przycisk bezpieczeństwa / E-Stop	Schneider E	8	Awaryjne zatrzymanie linii
21	Wyłącznik krańcowy / XCKN	Schneider E	10	Kontrola pozycji elementów
22	Elektrozawór pneumatyczny / VUVG	Festo	12	Sterowanie siłownikami
23	Siłownik pneumatyczny / ADN	Festo	12	Operacje montażowe
24	System transportowy / TS2plus	Bosch Rex	1	Transport detali między stanowiskami
25	Sterownik przenośn / FlexLink Drive	FlexLink	1	Sterowanie transportem

W tabeli 3 zestawiono najważniejsze elementy, które zostaną zastosowane na linii produkcyjnej. Na rysunki 31 pokazano ich rozmieszczenie, a na rys. 32 projekt hali produkcyjnej. Automatyzacja linii do produkcji samochodowych lamp LED będzie realizowana przez system sterowania składający się z:

- sterowników PLC – 1 szt., panele HMI – 2 szt.
- robotów przemysłowych – ok. 6 szt.
- systemów wizyjnych – 3 szt.
- czujników – ok. 80 szt. i napędów elektrycznych – ok. 12 szt.
- systemów bezpieczeństwa.

Rys. 31. Urządzenia zastosowane na linii



Rys. 32. Plan hali produkcyjnej (ChatGpt)

Centralnym elementem systemu sterowania jest sterownik PLC, który koordynuje pracę wszystkich urządzeń. Przykładowo można zastosować sterownik Siemens S7-1500 albo Allen-Bradley ControlLogix albo Mitsubishi iQ-R. Będzie on odpowiadał za: obsługę czujników i napędów, synchronizację pracy linii, nadzór nad robotami, kontrolę transportu. Zaplanowano także zastosowanie systemu HMI oraz systemu wizualizacji SCADA, który umożliwi: monitorowanie pracy linii, diagnostykę, rejestrację danych produkcyjnych oraz zarządzanie alarmami. Jako panel operatorski proponowane jest zastosowany zostanie Siemens TP1200

Comfort. Jako system SCADA możliwe będzie użycie systemu: InTouch albo WinCC. Będzie on połączony z PLC za pośrednictwem magistrali Ethernet.

Przewiduje się zastosowanie robotów przemysłowych, które będą wykonywać większość operacji montażowych tj. montaż radiatora, płytki PCB LED, optyki i aplikacja kleju oraz obsługa stanowisk testowych. Zaproponowano zastosowanie robotów 6-cio osiowych typu ABB IRB 1200 albo KUKA KR10 o parametrach: udźwig: 10 kg, zasięg: 900–1300 mm, dokładność: ±0,02 mm. Każdy z robotów ma własne sterowniki, które będą łączyły się ze sterownikami PLC za pośrednictwem magistrali Profinet. Głównymi sygnałami będą: Start cyklu, Zakończenie, Błąd.

Transport między stanowiskami odbywa się przy użyciu: przenośnika paletowego, buforów międzyoperacyjnych oraz systemu identyfikacji elementów RFID. Do transportu elementów pomiędzy stanowiskami roboczymi wykorzystane zostanie przenośnik paletowy. Oprócz tego zastosowane zostaną: przenośniki rolkowe, napędy elektryczne, moduły transferowe i bufory. Przykładowo mogą to być systemy transportowe Bosch Rexroth TS2plus albo FlexLink.

Na linii zastosowany zostanie zestaw czujników, służących do wykrywania elementów, kontroli położenia oraz monitorowania procesu. Planuje się użycie następujących rodzajów czujników:

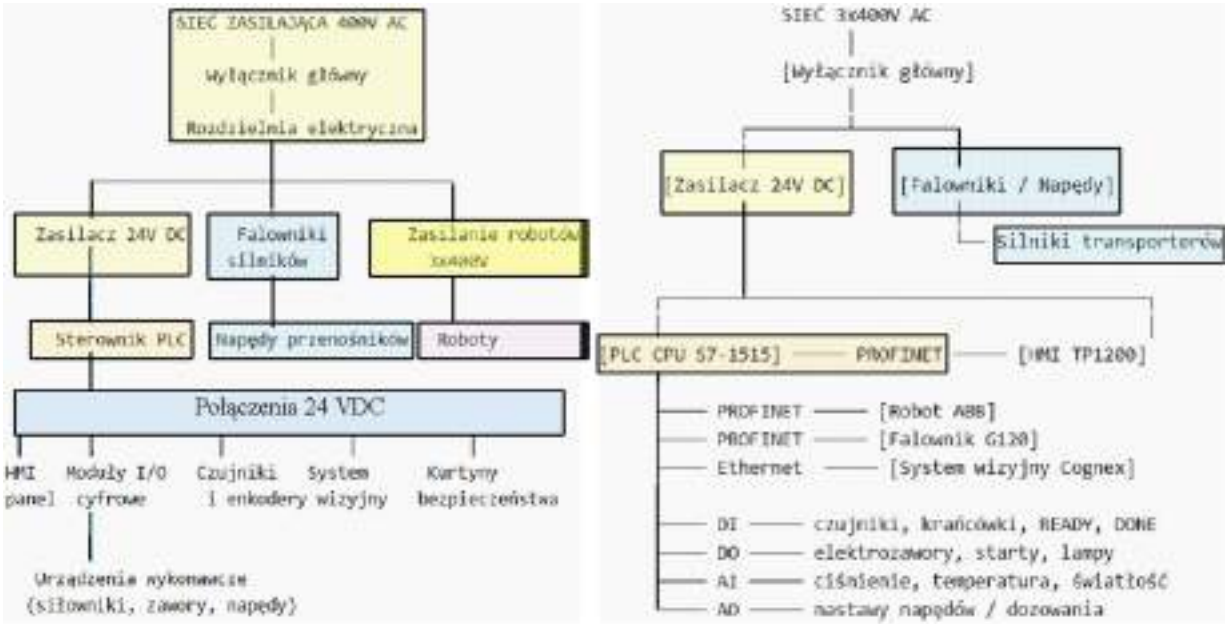
- czujniki indukcyjne, zastosowanie: detekcja elementów metalowych, rodzaj np.: IFM, Pepperl+Fuchs
- czujniki optyczne, zastosowanie: wykrywanie obecności detalu, kontrola pozycji; przykład: SICK, Omron
- czujniki wizyjne, zastosowanie: kontrola montażu, wykrywanie wad, np.: Keyence, Cognex
- system wizyjny, zastosowane do wykrywania i rozpoznawania elementów, kontroli pozycję elementów, poprawność montażu, obecność komponentów; przykładem jest Cognex In-Sight; parametry: rozdzielczość 2–5 MP, analiza obrazu w czasie rzeczywistym, komunikacja z PLC: Ethernet.

Oprócz czujników zastosowane zostaną także napędy elektryczne, głównie asynchroniczne, do napędzania przenośników i podajników. Koniecznym elementem linii produkcyjne będzie także system bezpieczeństwa, który zapewni ochronę operatorów za pomocą: kurtyn świetlnych, wyłączników bezpieczeństwa i stop oraz sterowników bezpieczeństwa. Dotyczy to

przede wszystkim kurtyn, zabezpieczających wejście człowieka w strefę pracy robota. Proponuje się zastosowanie kurtyn świetlnych typu SICK deTec oraz sterowników bezpieczeństwa typu Siemens S7-1500F. Nieodzownym elementem systemu sterowania jest zastosowanie systemu komunikacyjnego, który pozwoli na realizowanie przez sieć przemysłową. Zaplanowano zastosowanie protokołów:

- Profinet – komunikacja PLC i robotów
- Ethernet/IP – komunikacja urządzeń
- IO-Link – czujniki inteligentne

Połączenia robota z PLC będzie realizowane przez PROFINET, za pomocą sygnałów:
PLC → Robot: Start – rozpoczęcie cyklu; Reset – kasowanie błędu
Robot → PLC: Ready – robot gotowy; Busy – robot zajęty; Done – cykl zakończony.
Połączenie systemu wizyjnego z PLC będzie realizowane przez Ethernet, za pomocą sygnałów:
Trigger – wykonanie zdjęcia; OK – poprawny montaż; NOK – błąd montażu.



Rys. 33. Zasilanie systemu automatyki

Ze względu na dużą liczbę produkcji lamp, projektowana linia będzie miała charakter produkcji seryjnej i masowej. Jest to obecnie najczęściej stosowane rozwiązanie w przemyśle. Linie ta będzie miała także charakter produkcji ciągłej, w której produkcja i wyposażenie są przygotowane do pracy bez przerw technologicznych. W związku z tym konieczne jest zastosowanie specjalistycznych układów automatyki, sterowania i monitoringu. Linia produkcyjna będzie zasilana z sieci 3 x 400 VAC. Zostaną w niej zastosowane bezpieczniki i wyłączniki, które chronią układ przed przeciążeniami i zwarciami oraz zapewniają

bezpieczeństwo pracowników. Szafy sterownicze są kluczowymi elementami, w których montuje się podzespoły zasilania i automatyki przemysłowej. Ogólny schemat zasilania linii do produkcji lamp pokazano na rys. 33. Poniżej przedstawiono zestawienie elementy automatyki, które będą zainstalowane na linii.

Tabela 4. Zestawienie doboru elementów automatyki na linii produkcji lamp LED

Grupa	Zastosowanie	Przykładowy model
Tester szczelności	Kontrola jakości	Ateq / Cincinnati Test Systems
Tester optyczny	Kontrola jakości	stanowisko fotometryczne z czujnikiem
Tester elektryczny	Kontrola jakości	dedykowane stanowisko testu prądu
System wizyjny	kontrola końcowa	Keyence CV-X/Cognex + oświetlenie
Siłowniki pneumatyczne	Dociski, blokady, ruchy	Festo ADN / SMC CQ2
Przygotowanie powietrza	Filtr, reduktor, smarów	Festo MS / SMC AC
Przekaźniki bezpieczeństwa	bezpieczeństwo	Pilz PNOZmulti
Robot 1	montaż radiatora i PCB	ABB IRB 1200
Robot 2	montaż optyki	ABB IRB 1100 / KUKA AGILUS
Robot 3, 4	Elektronika, zgrzew.	ABB IRB 1200
Robot 5, 6	klejenie / manipulacja	ABB IRB 1300
Transport	system paletowy	Bosch Rexroth TS2plus
Falowniki+ przekładnie 2	napędy transporterów	SEW-Eurodriv albo SINAMICS G120
Serwonapędy, podawanie …	osie precyzyjne	Siemens SINAMICS S210
Czujniki indukcyjne	detekcja metalu	IFM IG, SMC D-M9 / Festo SME
Czujniki optyczne	detekcja obecności	SICK WL12
System wizyjny do montażu	kontrola montażu	Cognex In-Sight 7000
Kurtyny świetlne	bezpieczeństwo	Dobór: SICK deTec4
Pneumatyka	Zawory/czujniki	Festo VUVG/ IFM PN / WIKA
Bezpieczeństwo	kurtyny	SICK deTec4
Switch	sieć PROFINET	Siemens Scalance X
Zasilacz	24 V DC	Siemens SITOP
Szafa	obudowa sterownicza	Rittal VX25

Połączenia robota z PLC będzie realizowane przez sieć PROFINET, za pomocą sygnałów:

PLC → Robot: Start – rozpoczęcie cyklu; Reset – kasowanie błędu

Robot → PLC: Ready – robot gotowy; Busy – robot zajęty; Done – cykl zakończony.

Połączenie systemu wizyjnego z PLC będzie realizowane przez Ethernet, za pomocą sygnałów: Trigger – wykonanie zdjęcia; OK – poprawny montaż; NOK – błąd montażu.

Sekwencja pracy zautomatyzowanej linii jest następująca:

S0 – linia zatrzymana → START

S1 – transport palety do stacji montażu → czujnik pozycji

S2 – montaż radiatora i PCB LED → robot zakończył operację

S3 – montaż optyki → pozycjonowanie OK

S4 – montaż elektroniki sterującej → test elektryczny OK

S5 – klejenie / zgrzewanie klosza → proces zakończony

S6 – test szczelności lampy → test OK

S7 – test optyczny → parametry OK

S8 – transport do pakowania → produkt odebrany

S9 – koniec cyklu

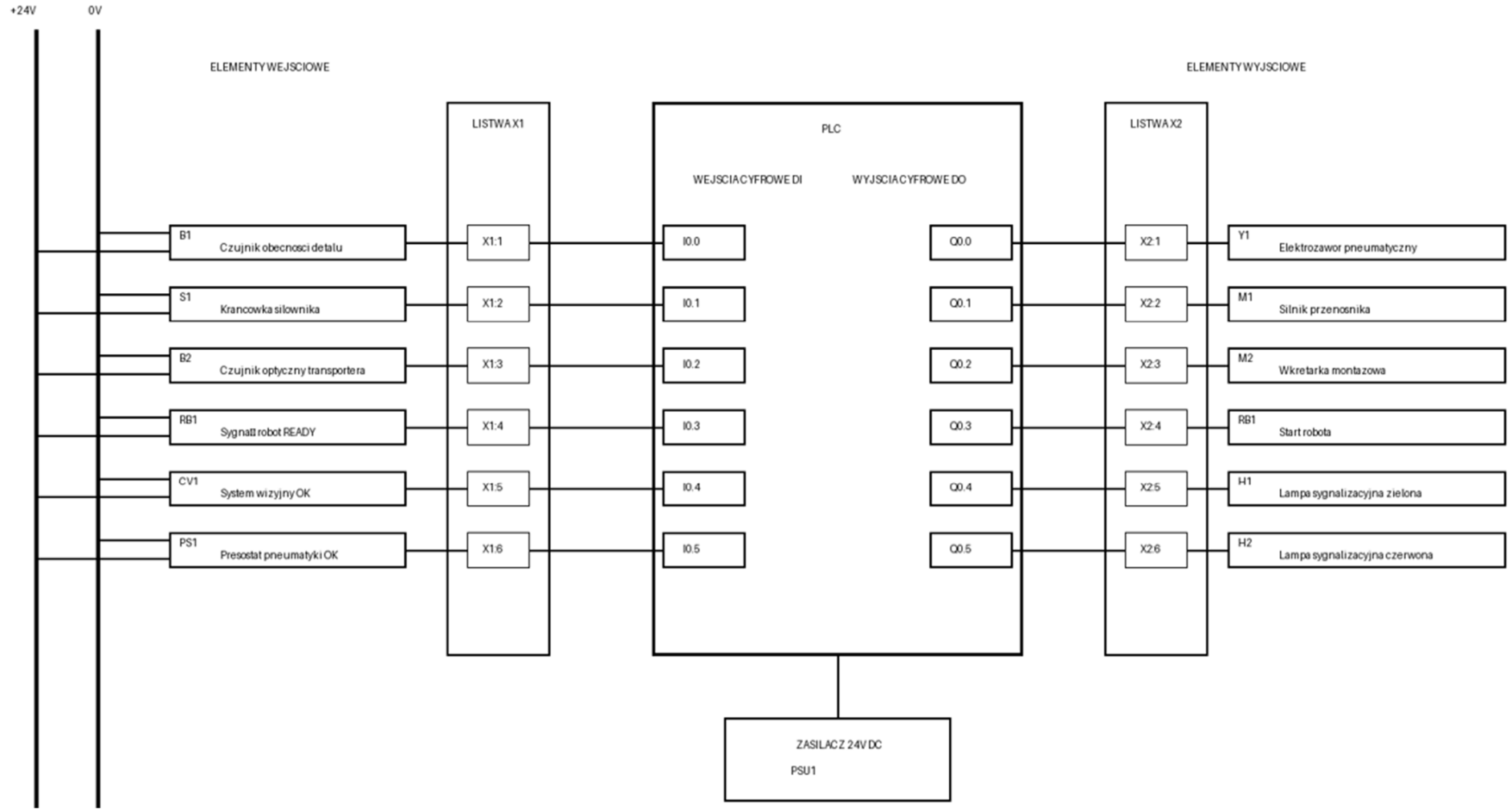
Tabela 5. Przykładowe adresy elementów we/wyj w sterowniku PLC na linii montażu lamp samochodowych

We. PLC	**Element**	**Funkcja**	**Wyj. PLC**	**Element**	**Funkcja**
I0.0	czujnik obecności detalu	wykrycie lampy	Q0.0	elektrozawór 1	sterowanie siłownikiem
I0.1	czujnik pozycji palety	pozycjonowanie	Q0.1	elektrozawór 2	sterowanie siłownikiem
I0.2	czujnik optyczny	kontrola transportu	Q0.2	start robota	sygnał cyklu
I0.3	krańcówka siłownika	pozycja wysunięta	Q0.3	reset robota	reset alarmu
I0.4	krańcówka siłownika	pozycja schowana	Q0.4	start przenośnika	transport
I0.5	sygnał robota READY	robot gotowy	Q0.5	stop przenośnika	zatrzymanie
I0.6	sygnał robota DONE	zakończenie cyklu	Q0.6	wkrętarka	montaż
I0.7	sygnał systemu wizyjnego	OK/NOK	Q0.7	lampa sygnalizac.	status linii
AI0	czujnik temp. radiatora	pom temp. radiatora			
AI1	czujnik ciśnienia	pom. ciśn. pneumat.			
AI1	czujnik natężenia światła	test lampy			

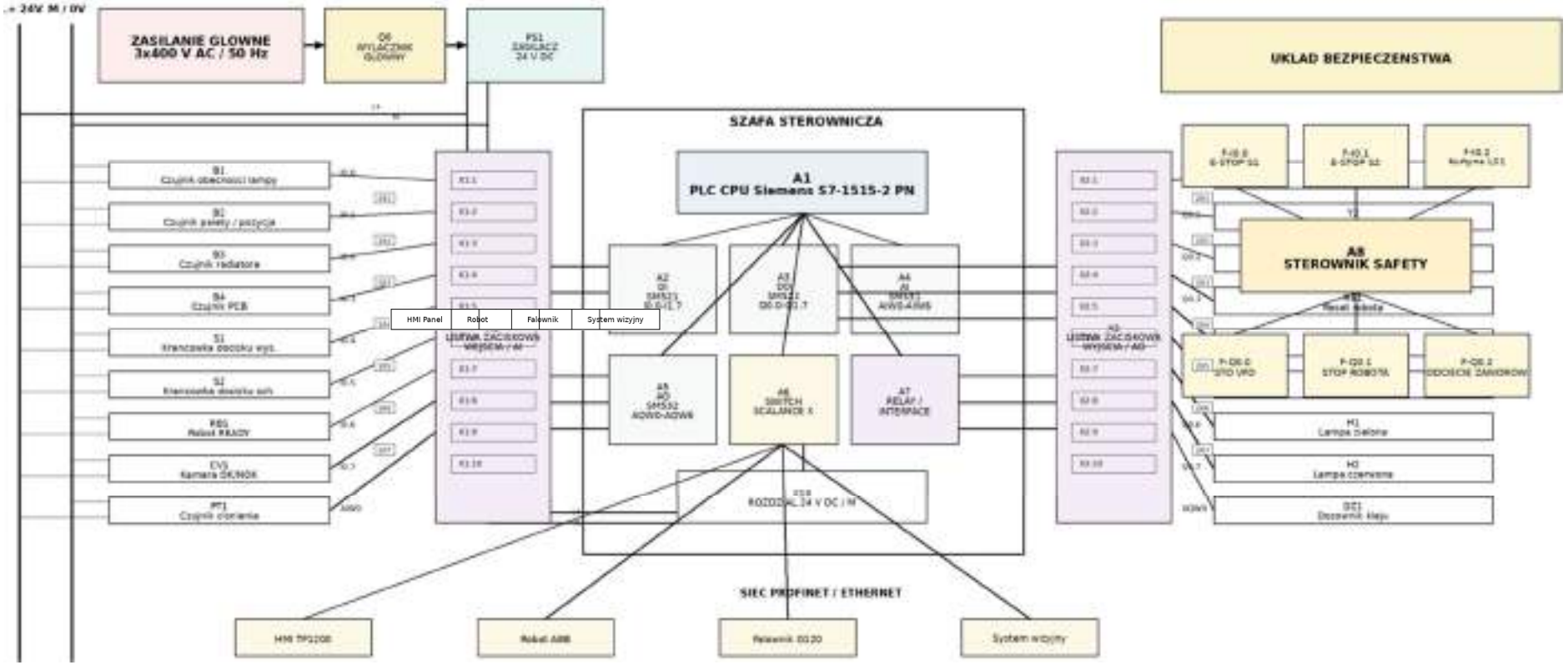
Poniżej, na rysunkach 34 i 35 pokazano zestawienia elementów automatyki i przykłady połączenia elementów elektrycznych i automatyki do wejść sterownika PLC.

Rys. 34. Ilustracja wybranych elementów automatyzacji na linii produkcji lampy LED

a)

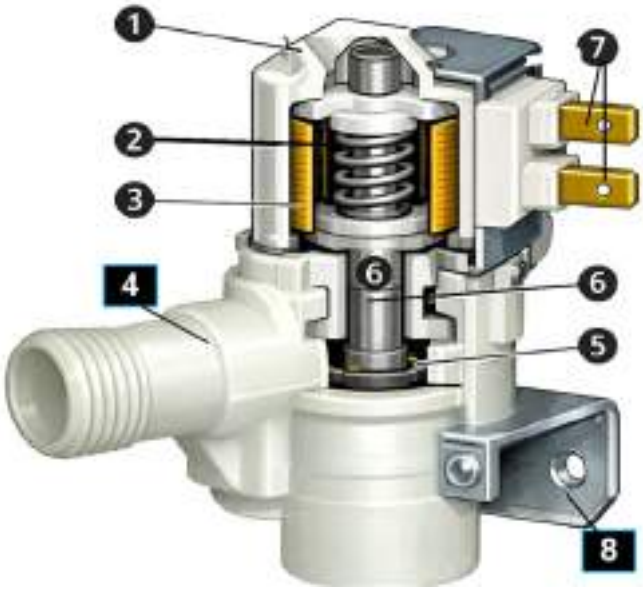

+24V
0V
ELEMENTY WEJSCIOWE
LISTWA X1
PLC
WEJSCIA CYFROWE DI
WYJSCIA CYFROWE DO
LISTWA X2
ELEMENTY WYJSCIOWE
B1
Czujnik obecnosci detalu
X1:1
I0.0
Q0.0
X2:1
Y1
Elektrozawor pneumatyczny
S1
Krancowka silownika
X1:2
I0.1
Q0.1
X2:2
M1
Silnik przenosnika
B2
Czujnik optyczny transportera
X1:3
I0.2
Q0.2
X2:3
M2
Wkretarka montazowa
RB1
Sygnatl robot READY
X1:4
I0.3
Q0.3
X2:4
RB1
Start robota
CV1
System wizyjny OK
X1:5
I0.4
Q0.4
X2:5
H1
Lampa sygnalizacyjna zielona
PS1
Presostat pneumatyki OK
X1:6
I0.5
Q0.5
X2:6
H2
Lampa sygnalizacyjna czerwona
ZASILACZ 24V DC
PSU1

b)



Rys. 35. Ilustracja podłączeń elementów do sterownika: a) na stanowisku montażu lampy LED, radiatora i płytki PCB, b) schemat blokowy według EPLAN całości połączeń systemu sterowania (wg ChatGPT)

Przykład 3: Koncepcja zautomatyzowanej linii do produkcji elektrozaworów pralek



Rys. 36. Budowa elektrozaworu (wg ChatGPT)

Celem projektu jest opracowanie koncepcji nowoczesnej, zrobotyzowanej linii do produkcji elektrozaworów stosowanych w pralkach automatycznych. Linia ma zapewnić wydajność 300 000 szt./rok, przy ograniczeniu udziału ludzi i powtarzalność montażu i testowania w zakresie kontroli szczelności i parametrów elektrycznych 100% wyrobów. Elektrozawory służą do włączania dopływu wody do bębna pralki. Będą zasilane napięciem sieciowym 230 V AC/50 Hz. Biorąc pod uwagę założoną wydajność oraz efektywność 85% i pracę 2-zmianową, można wyznaczyć czas pracy 12 240 000 s/ na rok, co daje czas taktu pracy 41 s.

Na rysunku 36 pokazano elektrozawór. Składa się on z elementów:

1) korpus z tworzywa
2) sprężyna
3) cewka
4) obudowa zaworu + przyłącze wody
5) membrana
6) rdzeń ruchomy
7) przyłącza elektryczne
8) element do mocowania

Produkcja będzie miała charakter wielkoseryjny – masowy z taktem docelowym: 35–40 s/szt.

Linia montażowa będzie składała się ze stacji:

STACJA 1 – Załadunek korpusu zaworu (4): podajnik wibracyjny, robot pick&place, Czujniki do kontroli obecności

STACJA 2 – Montaż membrany (5): podajnik elastycznych elementów, manipulator, kontrola położenia

STACJA 3 – Montaż rdzenia (6): podawanie rdzenia stalowego, precyzyjne osadzenie, czujnik pozycji

STACJA 4 – Montaż sprężyny (2): podajnik sprężyn, automatyczne osadzanie, kontrola

STACJA 5 – Montaż cewki (3): podajnik cewek, robot SCARA - nasunięcie na rdzeń

STACJA 6 – Montaż obudowy cewki (1), docisk, kontrola montażu

STACJA 7 – Montaż przyłączy elektrycznych (7): wcisk konektorów, połączenie z cewką, kontrola poprawności (kamera)

STACJA 8 – Montaż elementu mocującego (8), podajnik blach, zaginania, wiercenia, robot montujący, wcisk, kontrola jakości

STACJA 9 – Kontrola wizyjna: obecność wszystkich elementów, poprawność montażu

STACJA 10 – Test elektryczny, rezystancja cewki

STACJA 11 – Test funkcjonalny, podanie napięcia, sprawdzenie ruchu rdzenia

STACJA 12 – Test szczelności, podłączenie, podanie powietrz, pomiar spadku ciśnienia

STACJA 13 – znakowanie (laser / etykieta), pakowanie.

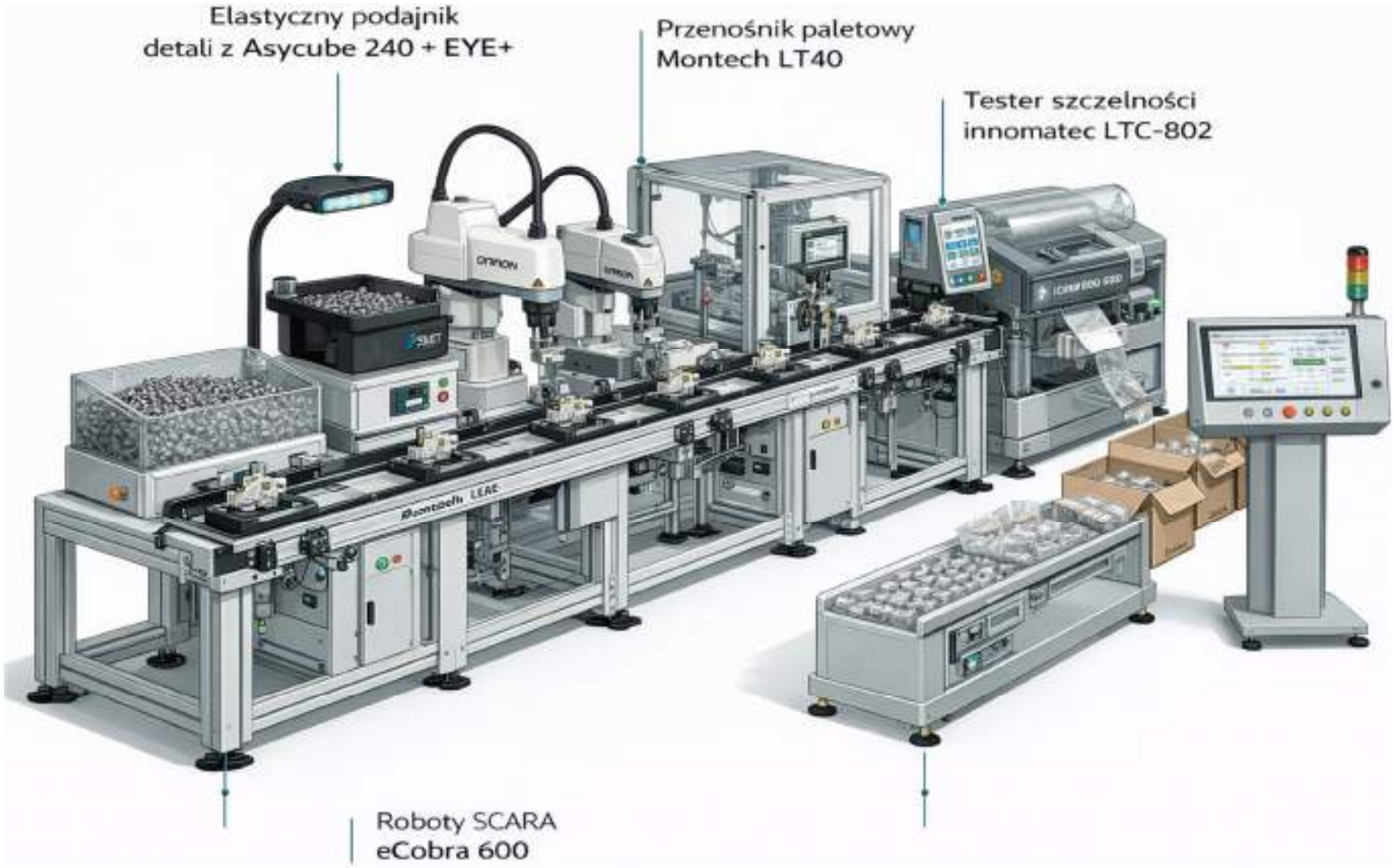
Do budowy zautomatyzowanej linii montażu elektrozaworu konieczne jest zastosowanie systemu transportu paletowego, automatycznych podajników elementów, manipulatorów montażowych, gniazd pozycjonująco-montażowych, stacji kontroli jakości, stanowisk testu elektrycznego i szczelności, układu znakowania, pakowania, centralnego systemu sterowania PLC oraz kompletnego systemu bezpieczeństwa i zasilania w media.

Tabela 6. Moduły i urządzenia na linii produkcyjnej

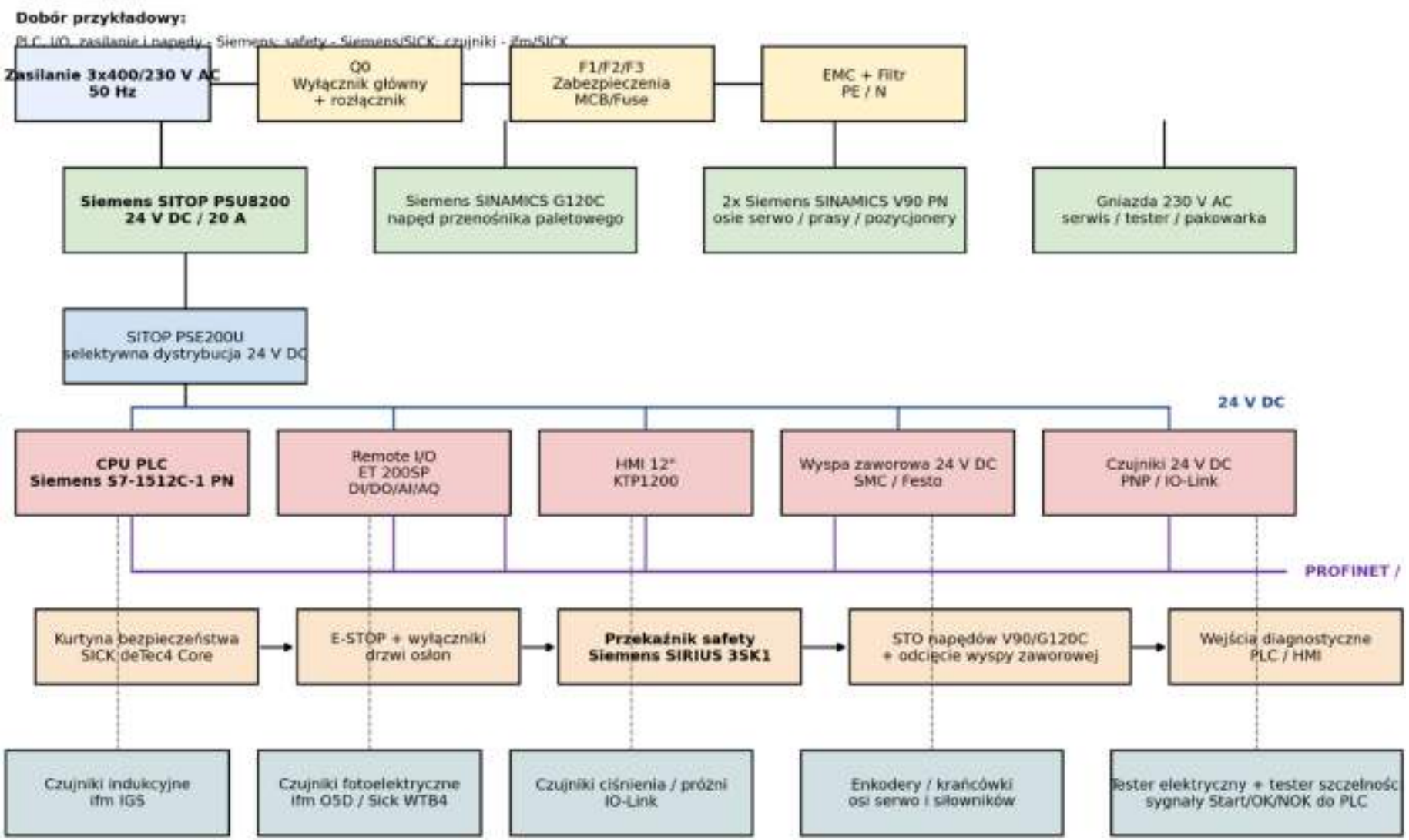
Urządzenie	Zastosowanie w zautomatyzowanej linii produkcyjnej
Transport paletowy Montech LT40	LT40 to systemem do transportu nośników detalu do stacji montażowych, prędkość do 20 m/min i udźwig nośnika do 16,8 kg.
Modułowy transfer system Bosch Rexroth TS 2plus	TS 2plus: modułowy system transferowy dla palet do 240 kg, z gotowymi modułami pozycjonowania i sterowania. Rozwiązanie raczej z zapasem, ale bardzo wiarygodne projektowo.
Podajnik Asyril Asycube 240 + Asyfill + EYE+	Do małych: sprężyn, konektorów, rdzeni, małych blaszek. Asycube 240 jest przeznaczony dla części ok. 5–40 mm. Zespół Asyfill + EYE+ odpowiada za dozowanie, separację i lokalizację części dla robota.
Podajnik RNA Automation bowl feeder	Dla części np. sprężyny albo konektora, feeder wibracyjny. Można też: zastosować elastyczny feeder dla części zmiennych i misowy dla części najbardziej powtarzalnych.
Robot OMRON eCobra 600 / 800	SCARA: zasięg 600 lub 800 mm i udźwig do 5,5 kg. Dla montażu cewki, rdzenia, membrany i blaszki mocującej to bardzo trafny wybór.
Robot 6-osiowy ABB IRB 1300	IRB 1300 udźwig 7–12 kg, może obsługiwać kilka stacji albo wykonywać bardziej złożone ruchy montażowe.
Chwytak elektryczny SCHUNK EGU	EGU ma szerokie opcje komunikacji przemysłowej i jest dobry do elastycznej produkcji wielowariantowej. Sprawdzi się przy odkładaniu korpusu, chwytaniu cewki lub elementu mocującego.
Chwytak DESTACO RP	Prosty, lekki 2-szczękowy chwytak pneumatyczny do szybkiej, powtarzalnej manipulacji.
Szybkowymienne bazy pod gniazda SCHUNK VERO-S NSE mini	Same gniazda muszą być wykonane specjalnie pod część, ale można zastosować system do szybkiej wymiany: samohamowny i otwierany pneumatycznie przy 6 bar, więc dobrze nadaje się jako baza pod wymienne przyrządy montażowe.
Tester cewki: Marposs E.D.C. / AMT5-W	Stacja EOL do testów elektrycznych i funkcjonalnych komponentów oraz automatycznych stanowisk EOL, mierząca rezystancję cewki, ciągłość połączeń i ewentualnie próbę zasilenia zaworu.
Tester uzwojeń / cewek Marposs AST5/W	Jest przeznaczony do pełnych testów uzwojeń, cewek i stojanów w produkcji. W Twojej linii można go potraktować jako gotową bazę do testu cewki przed montażem albo na końcu linii.
Tester szczelności innomatec LTC-802	LTC-802 z integracją do automatyki do zautomatyzowanej linii: obsługuje wiele metod testu, ma szeroką komunikację przemysłową, integrację z PLC i raportowanie danych.
Tester szczelności innomatec LTC-503	LTC-503 entry-level nadaje się do standardowych testów szczelności, oferuje kilka metod pomiaru i jest łatwy do integracji tester wejściowy. Jego zastosowanie będzie wymagać wprowadzenia modyfikacji.
innomatec DF-10 Leak Simulator	Do linii testowej warto dodać też wzorzec przecieku Wzorzec przecieku / symulator do walidacji stanowiska.

Urządzenie	Zastosowanie w zautomatyzowanej linii produkcyjnej
Automatyczna pakowarka woreczkowa AUTOBAG 500	Autobag 500 to automatyczna maszyna do napełniania i zgrzewania woreczków, z wydajnością ponad 100 worków/min. Dla pojedynczego elektrozaworu lub kompletu serwisowego to bardzo praktyczne rozwiązanie końca linii.
Głowica etykietująca HERMA 500	HERMA 500 jest rdzeniem przemysłowych maszyn etykietujących HERMA i nadaje się do nanoszenia etykiety z kodem partii, datą i numerem wyrobu na opakowanie lub tackę.
Druk etykiety cab HERMES Q	HERMES Q jest lepszy niż osobna drukarka plus aplikator.

Najważniejsze elementy automatyzacji są następujące: przenośnik paletowy, roboty SCARA, podajniki elastyczne, system wizyjny, test szczelności, PLC + HMI, system bezpieczeństwa. Poniżej, na rysunku 36, przedstawiono uproszczony rysunek linii produkcyjnej.



Rys. 37. Budowa linii produkcyjnej (wg ChatGPT)



Rys. 38. Schemat układu sterowania linią montażu elektrozaworu (wg ChatGPT)

Schemat elektryczny (rys. 38) przedstawia układ, na którym wyszczególniono blok zasilania 3x400/230 V AC oraz połączenia wszystkich najważniejszych urządzeń i elementów.

Tabela 7. Elementy systemu sterowania

Obszar	Dobór
Sterownik główny PLC	Siemens SIMATIC S7-1512C-1 PN
Zdalne I/O	Siemens ET 200SP
Panel operatorski	Siemens KTP1200 Comfort
Zasilacz 24 V DC	Siemens SITOP PSU8200 24 V / 20 A
Selektywna dystrybucja 24 V	Siemens SITOP PSE200U
Napęd przenośnika paletowego	Siemens SINAMICS G120C
Napędy osi pozycjonujących / serwo	Siemens SINAMICS V90 PN + silnik S-1FL6
Safety	Siemens SIRIUS 3SK1
Kurtyna bezpieczeństwa	SICK deTec4 Core
Czujniki indukcyjne	ifm – indukcyjne IGS / odpowiednik M12/M18 PNP
Czujniki fotoelektryczne	ifm O5D lub SICK WTB4
Czujniki ciśnienia / próżni	IO-Link, 24 V DC
Wyspa zaworowa pneumatyki	SMC lub Festo, 24 V DC

Do zasilania układu automatyki zastosowano SITOP PSU8200. Do rozdziału i ochrony gałęzi zasilania 24 V DC, zaproponowano zastosowanie SITOP PSE200U, który jest tzw. modułem selektywności elektronicznej dla obwodów 24 V DC. Sterownik PLC typu S7-1512C-1 PN nadaje się do zwartej, średniej linii montażowej. Moduł ET 200SP to rozszerzenie I/O, w którym zastosowano typowe moduły DI 16x24 V DC i DQ 16x24 V DC/0,5 A, które pozwalają podłączyć czujniki, elektrozawory i sygnały pomocnicze. Do napędzania przenośnika paletowego dobrano napęd typu SINAMICS G120C.Firma Siemens opisuje go jako kompaktowy przemiennik do aplikacji takich jak przenośniki. Dla napędzania osi pozycjonujących i małych napędów montażowych, wybrano SINAMICS V90 PN, który może komunikować się do sieci PROFINET i współpracuje z Motion Control S7-1500.Do detekcji

pozycji i obecności części metalowych różnych elementów na linii produkcyjnej przyjęto zastosowanie czujników indukcyjnych firmy ifm. Do detekcji innych komponentów wykonanych z metali nieżelaznych albo z tworzywa planuje się zastosowanie czujników fotoelektrycznych, ewentualnie z tłumieniem tła. Firma ifm ma szeroką ofertę czujników indukcyjnych oraz fotoelektrycznych. Seria O5D jest przeznaczona do niezależnej od koloru detekcji z tłumieniem tła.

Poniżej pokazano schemat adresacji I/O dla stacji S1 – Załadunek korpusu.

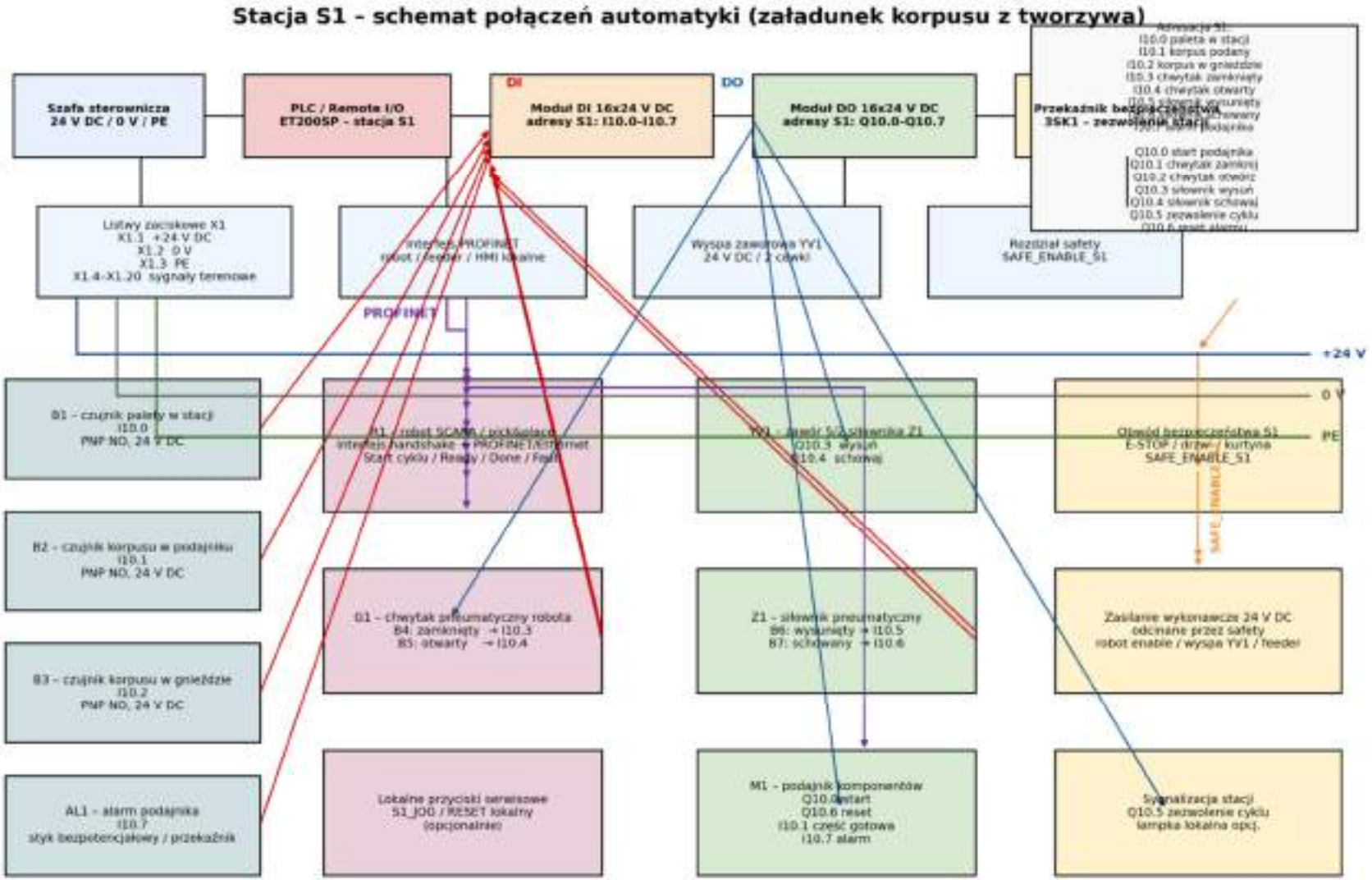
Adres	Typ	Nazwa sygnału	Opis
I10.0	DI	S1_Paleta_w_stacji	Czujnik obecności palety
I10.1	DI	S1_Korpus_podany	Korpus dostępny w podajniku
I10.2	DI	S1_Korpus_w_gniezdzie	Potwierdzenie odłożenia detalu
I10.3	DI	S1_Chwytak_zamkniety	Potwierdzenie chwytaka
I10.4	DI	S1_Chwytak_otwarty	Potwierdzenie chwytaka
I10.5	DI	S1_Silownik_wysuniety	Potwierdzenie siłownika
I10.6	DI	S1_Silownik_schowany	Potwierdzenie siłownika
I10.7	DI	S1_Alarm_podajnika	Błąd podawania

Adres	Typ	Nazwa sygnału	Opis
Q10.0	DO	S1_Start_podajnika	Start podajnika korpusu
Q10.1	DO	S1_Chwytak_zamknij	Sterowanie chwytakiem
Q10.2	DO	S1_Chwytak_otworz	Sterowanie chwytakiem
Q10.3	DO	S1_Silownik_wysun	Ruch roboczy
Q10.4	DO	S1_Silownik_schowaj	Powrót
Q10.5	DO	S1_Zezwolenie_cykl	Start cyklu stacji
Q10.6	DO	S1_Reset_alarmu	Reset podajnika
Q10.7	DO	Rezerwa_S1	Rezerwa

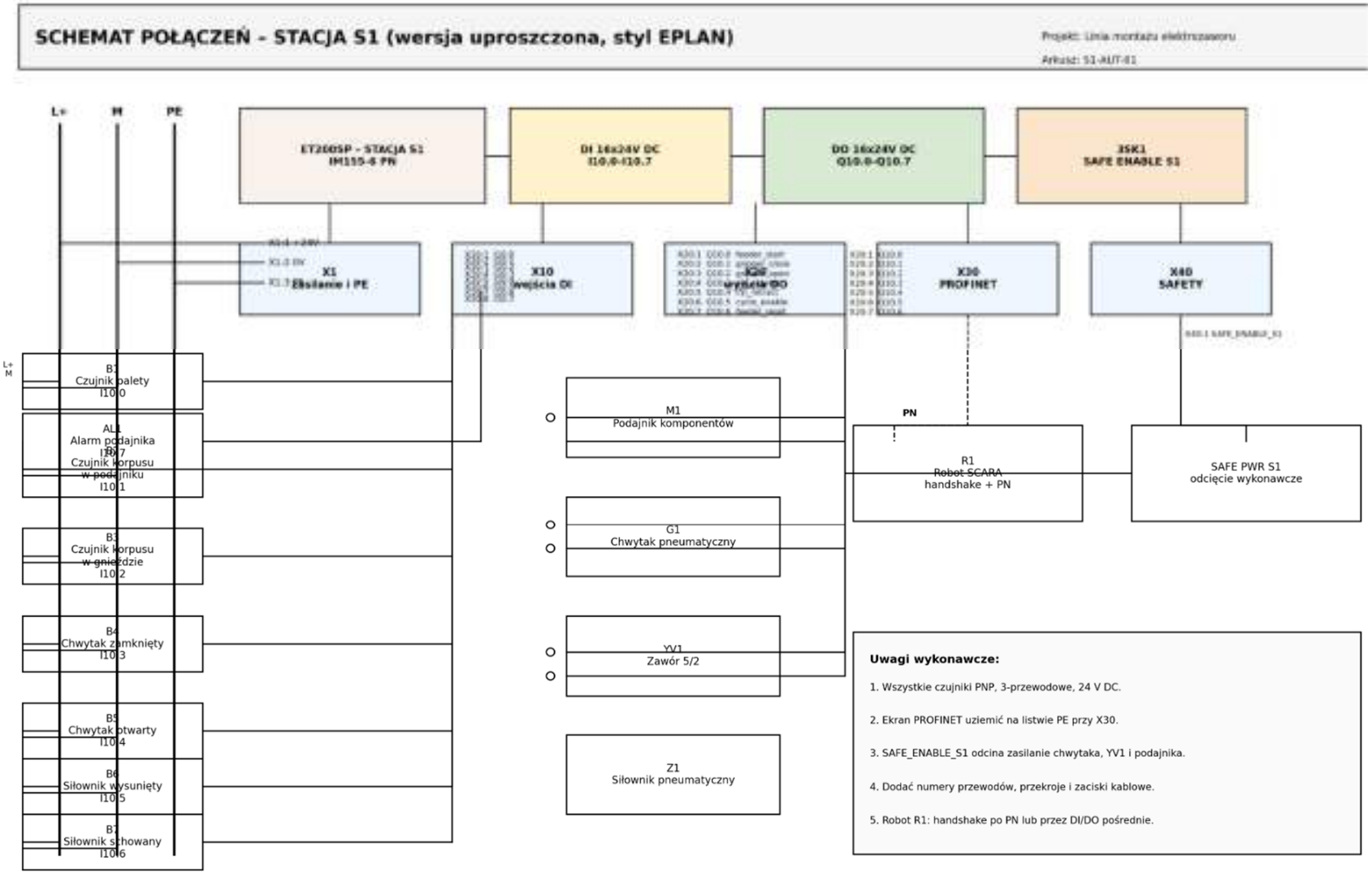
Plik PNG: Schemat połączeń stacji S1

Na schemacie są pokazane między innymi:

- szafa sterownicza 24 V DC / 0 V / PE,
- PLC / ET200SP z modułami **DI** (adresy **I10.0–I10.7)** i **DO** (adresy **Q10.0–Q10.7**,
- listwy zaciskowe, interfejs **PROFINET**,
- podajnik komponentów,
- robot pick-and-place, chwytak pneumatyczny, zawór 5/2 i siłownik pneumatyczny,
- obwód bezpieczeństwa z sygnałem **SAFE_ENABLE_S1**.



Rys. 39. Schemat ogólny połączeń elementów automatyki – stacja S1 (wg ChatGPT)



Rys. 40. Schemat układu połączeń elementów automatyki do sterownika PLC – stacja S1, z wyszczególnieniem list połączeń (wg ChatGPT)

Podsumowanie

Projektowanie zautomatyzowanych linii produkcyjnych wymaga od osoby wykonującej posiadania rozległej i pogłębionej wiedzy na temat budowy urządzeń, technologii ich wykonania i produkcji, maszyn i urządzeń zarówno do obróbki metali, jak i do przetwarzania tworzyw sztucznych, a także substancji chemicznych, w tym spożywczych i innych. Projektant powinien mieć także wiedzę na temat transportu wewnątrz zakładowego oraz zarządzania produkcją. Bardzo ważna jest także znajomość automatyki, w zakresie jej podstaw oraz elementów i układów automatyzacji. Szczególnie ważna jest wiedza dotycząca sterowników przemysłowych i umiejętność ich programowania. Na prawie wszystkich zautomatyzowanych liniach produkcyjnych powszechnie stosuje się roboty przemysłowe, podajniki, transportery, których znajomość jest również niezbędna. Dlatego w czasach obecnych, do przygotowania koncepcji linii zautomatyzowanej, sensownym jest skorzystanie z systemów, bazujących na metodach sztucznej inteligencji w postaci programów ChatGPT, Copilot i Gemini i innych. W niniejszym opracowaniu z sukcesem wykorzystano te właśnie narzędzia. Trzeba jednak zauważyć, że wygenerowane przez nie rozwiązania miały charakter bardzo wstępny i ogólnikowy. Pokazały jednak dość dokładnie problemy i złożoność opracowanych linii zautomatyzowanych. Można łatwo zauważyć, że wymienione wyżej systemy, kompletnie nie „rozumieją” rysunków produkowanych urządzeń, i w związku z tym wykonane projekty zawierają niedokładności a nawet błędy, które projektant, musi wykryć i poprawić. Użycie np. ChatGPT może znacząco przyspieszyć i usprawnić projektowanie.

Bibliografia

Domińczuk J., Kost G., Łebkowski P., Automatyzacja i robotyzacja procesów produkcyjnych, PWE, 2021.

Gola A., Zając J., Kost G., Integracja systemów wytwarzania, 2020.

Jarzębowska E., Dynamika i sterowanie układami mechanicznymi, PWN, 2021.

Kasprzyk J., Hajda J., Programowanie sterowników PLC, Wydawnictwo Pracowni Komputerowej Jacka Skalmierskiego, 1998.

Mikulczyński T., Samsonowicz Z., Więcławek R., Automatyzacja procesów produkcyjnych, 2015.

Podsiadły M., Automatyka przemysłowa dla początkujących, Helion, 2025.

Poradnik mechatronika, REA.

Szelerski M. W., Automatyka przemysłowa w praktyce, Projektowanie, modernizacja i naprawa, 2016, Wydawca: KaBe.

Szelerski M.W., Robotyka przemysłowa, KaBe, 2021.

Urządzenia i systemy mechatroniczne, REA.

Fundusze Europejskie dla Wielkopolski
Dofinansowane przez Unię Europejską
SAMORZĄD WOJEWÓDZTWA WIELKOPOLSKIEGO

Rachunek kosztów w ujęciu inżynierskim

dr Bronisław Bryl

Wstęp

Funkcjonowaniu w gospodarce rynkowej przedsiębiorstwom towarzyszy wysoka konkurencja i zmieniające się warunki otoczenia gospodarczego. Ich rozwój zależy w głównej mierze od szybkiego i elastycznego dostosowania się do zmian otocznia i rosnących oczekiwań klientów. Przedsiębiorstwa są zmuszone do oferowania produktów i świadczonych usług wysokiej jakości przy relatywnie niskim koszcie. Aby było to możliwe, decydenci zmuszeni są do ciągłego podejmowania różnorodnych decyzji ekonomicznych. Natomiast aby były skuteczne i efektywne, wymagają wspomagania odpowiednimi narzędziami. Narzędzia te muszą dostarczać niezbędnych do tego informacji (A. Karmańska, 2022).

Systemem, który dostarcza niezbędnych informacji do efektywnej i skutecznej realizacji procesów gospodarczych jest system rachunkowości (P. Szczypa, 2017). Rachunkowość jest określana systemem informacyjnym przedsiębiorstwa, dostarczającym informacji niezbędnych do podejmowania szybkich i racjonalnych decyzji, dotyczących planowania, organizowania, motywowania i kontrolowania prowadzonej działalności.
Informacje te kierowane są do szerokiego grona interesariuszy, którzy niekoniecznie muszą być ekonomistami lub nawet księgowymi. Funkcjonowanie w otoczeniu gospodarki rynkowej wymusza posiadanie podstawowej wiedzy ekonomicznej i finansowej, a zwłaszcza tej dotyczącej kosztów. Dotyczy to osób posiadających przygotowanie techniczne, w tym specjalistów z zakresu automatyki i robotyki. Niekiedy można mieć z pozoru bardzo dobry pomysł na rozwiązanie danego zagadnienia lub problemu, ale ograniczenia finansowe, w tym generowane koszty, spowodują jego nierealność w realizacji. Z tego też względu określony zakres wiedzy dotyczący rachunku kosztów staje się niezbędny nawet dla tej grupy zawodowej.

Dość często zdarza się, że osoby z przygotowaniem technicznym zajmują stanowiska kierownicze średniego i wyższego szczebla. Praca w tych obszarach wielokrotne związana jest podejmowaniem zróżnicowanych decyzji, których podjęcie wymaga wsparcia informacjami finansowymi. Ranga tych decyzji uzależniona jest od szczebla stanowiska kierowniczego. W przypadku stanowisk wyższego szczebla (prezes, członek zarządu czy

dyrektor pionu), decyzje te mogą dotyczyć realizacji, kontynuacji bądź zaniechani prowadzenia określonego projektu. Na stanowiskach kierowniczych średniego szczebla decyzyjność może dotyczyć chociażby merytorycznej akceptacji kosztów obciążających dany dział.

Opracowanie i przygotowanie projektu w obszarze automatyki i robotyki, obok zagadnień technicznych (rysunki, rzuty, obliczenia) może wymagać kosztorysu, który będzie określał wartości kosztów związanych z jego realizacją. Kosztorys wykonawczy jest wtedy integralną częścią opracowanego projektu i stanowi finansowe założenia jego realizacji.

Z taką sytuacją prawie zawsze można zetknąć się w przypadku przygotowania oferty dla potencjalnego nabywcy, który poszukuje oferentów poprzez ogłoszony przetargu w trybie zamówień publicznych. Skalkulowana i zaoferowana cena opiera się na przewidywanych kosztach realizacji potencjalnego zlecenia. Obok informacji technicznych należy tam podać również dane finansowe, będące elementem składanej oferty.

Przytoczone powyżej najczęściej występujące w praktyce sytuacje potwierdzają potrzebę posiadania przez osoby zajmujące się zagadnieniami technicznymi, wiedzy o kosztach, ich podziale czy sposobie rozliczania. Nie są to zagadnienia pozostawione tylko dla ekonomistów, analityków i księgowych. Specjaliści z zakresu automatyki i robotyki muszą mieć podstawową wiedzę i orientować się w obszarze rachunku kosztów.

1. Rachunek kosztów

Prowadzenie działalności gospodarczej jest immanentnie związane z ponoszeniem kosztów, których nie da się nie ponosić, ale należy je minimalizować. W literaturze rachunek kosztów jest definiowany w zróżnicowany sposób, a definicje te różnią się pojemnością i zakresem.

Według S. Sojaka, rachunek kosztów jest systemem identyfikacji, pomiaru, przetwarzania informacji o kosztach dla celów rachunkowości finansowej (według wymogów sprawozdawczych) i rachunkowości zarządczej (według potrzeb procesów zarządzania) (S. Sojak, 2000). Z kolei przez A. Jarugę, W. Nowak, A. Szychtę**,** rachunek kosztów jest określany jako system polegający na badaniu i transformowaniu informacji o kosztach działań przeszłych, bieżących i zamierzonych, według przyjętego modelu, w celu wspomagania zarządzania przedsiębiorstwem. (A. Jaruga, W. Nowak, A. Szychta, 2001). Najbardziej syntetycznie rachunek kosztów definiuje E. Nowak. Stwierdza on, że jest to proces ustalania kosztów prowadzenia działalności gospodarczej polegającej na wytwarzaniu i sprzedaży

wyrobów i świadczeniu usług (E. Nowak, 2001). Jest to najkrótsza definicja, ale jednocześnie najbardziej przyswajalna, zwłaszcza dla tych, którzy nie są ekonomistami.

Przedstawiając istotę rachunkowości jako systemu, Z. Kołaczyk podkreśla, że w wielu opracowaniach zagranicznych i krajowych wyodrębnia się w ramach rachunkowości jej części składowe, a mianowicie:

- księgowość, rozumianą jako proces rejestracji zdarzeń gospodarczych,
- kalkulację (rachunek kosztów), rozumianą jako ustalenie kosztu jednostkowego produktów na podstawie zarejestrowanych operacji gospodarczych o charakterze kosztowym,
- sprawozdawczość finansową, rozumianą jako agregację operacji gospodarczych o charakterze majątkowym i wynikowym, stanowiącą produkt obu poprzednich czynności (Z. Kołaczyk, 1996).

Powyższa struktura systemu rachunkowości jest określana mianem tradycyjnej. Należy zauważyć, że pomimo ramowego określenia poszczególnych części składowych rachunkowości, stanowi ona kompleksowe jej ujęcie, gdyż zawiera w elementy rachunkowości finansowej zarządczej.

Źródło: E. Nowak, 2018.

2. Koszt oraz pojęcia bliskoznaczne

Koszt jest nieodłącznym elementem związanym z prowadzeniem działalności gospodarczej, gdyż wytwarzanie wyrobów gotowych lub świadczenie usług wymaga wykorzystania różnych środków gospodarczych, co znajduje wyraz w kosztach.

Koszt jest kategorią ekonomiczną, która oznacza wyrażone w pieniądzu wartości pracy żywej oraz zasobów majątkowych zużytych w danym okresie w celu wytworzenia

wyrobów świadczenia usług i wykonywania określonych funkcji (E. Nowak 2001). Analizując pojęcie kosztu należy zwrócić uwagę na następujące jego cechy:

- celowe zużycie,
- poniesione na skutek prowadzonej działalności gospodarczej (związane z nią),
- koszt jest przyporządkowany do danego okresu,
- jest wyrażony w mierniku pieniężnym.

Obok pojęcia kosztu, funkcjonują pojęcia bliskoznaczne, które mimo podobieństw wyrażają inne znaczenie ekonomiczne. Ich odróżnienie jest istotne między innymi w celu prawidłowego konstruowania rachunku kosztów, a ich brak odróżniania może skutkować podjęciem błędnych decyzji. Do tych pojęć należy zaliczyć: nakład, wydatek i stratę nadzwyczajną.

Nakład jest kategorią ekonomiczną oznaczającą wyrażone w jednostkach naturalnych zużycie siły roboczej oraz zasobów majątkowych przedsiębiorstwa. Charakterystyczną cechą nakładu jest to, iż wyrażony jest w jednostkach naturalnych – kg, szt., rbh, itd.

Wydatek – jest zmniejszeniem stanu środków pieniężnych w kasie lub na rachunku bankowym jednostki gospodarczej.

Strata nadzwyczajna – to strat powstała na skutek zdarzeń trudnych do przewidzenia, poza działalnością operacyjną jednostki i niezwiązana z ogólnym ryzykiem jej prowadzenia.

Spośród analizowanych pojęć najczęściej kojarzone ze sobą są pojęcia kosztu i wydatku.

Przykład 1

Przeczytaj uważnie podane informacje dotyczące przedsiębiorstwa produkcyjnego i określ w kolumnie „Odpowiedź", czy wyraża ona: nakład (N), koszt (K), stratę nadzwyczajną (SN), czy wydatek środków pieniężnych (W).

Lp.	Informacja	Odpowiedź
1.	Zakup za gotówkę (300 zł) materiałów, które będą wykorzystane w procesie produkcji	
2.	Zużycie w procesie produkcji 40 litrów materiałów	
3.	Wartość zużytych w procesie produkcji materiałów (200 zł)	
4.	Wartość materiałów zniszczonych w wyniku pożaru magazynu (100 zł)	

5.	Zużycie urządzeń produkcyjnych (3 000 zł)	
6.	Czas pracy urządzeń produkcyjnych 10 maszynogodzin/ dzień	
7.	Pracownicy produkcyjni przepracowali łącznie 120 roboczogodzin	
8.	Naliczono wynagrodzenia brutto pracowników produkcyjnych (14 000 zł)	
9.	Przelano na konta osobiste wynagrodzenia pracowników produkcyjnych (9 500 zł)	
10.	Wartość skradzionych wyrobów gotowych (600 zł)	

Odpowiedzi

Nakład (N): 2, 6, 7,

Koszt (K): 1, 3, 5, 8,

Wydatek (W): 1, 9,

Strata nadzwyczajna (SN): 4, 10,

3. Podział kosztów

Ponoszone przez przedsiębiorstwo prowadzące działalność gospodarczą koszty, mogą być klasyfikowane według różnych kryteriów. W zależności od przyjętych kryteriów otrzymuje się odpowiednie przekroje klasyfikacyjne kosztów. Do podstawowych przekrojów klasyfikacyjnych kosztów, należy podział kosztów według:

- typów i rodzajów działalności,
- rodzajów,
- sposobów odniesienia kosztów na produkty,
- stopnia zależności od wielkości produkcji,
- struktury wewnętrznej kosztów,
- istotności kosztów przy podejmowaniu decyzji.

Powyższe klasyfikacje nie wyczerpują wszystkich możliwości ich podziału. Zaprezentowane tylko te, z którymi najczęściej można się spotkać i są najpowszechniej stosowane. Przedsiębiorstwa mogą stosować różnorodne podziały kosztów, które są pomocne w procesie decyzyjnym (J. Matuszewicz, 2002).

3.1. Koszty według rodzajów działalności

Dokonując podziału kosztów według typów działalności (miejsc ich powstawania), można uzyskać informacje o miejscu ich poniesienia i powstania. Można wyodrębnić następujące typy działalności:

- koszty podstawowej działalności operacyjnej,
- koszty pozostałej działalności operacyjnej,
- koszty działalności finansowej.

Spośród kosztów działalności operacyjnej można wyodrębnić koszty:

- działalności podstawowej,
- działalności pomocniczej,
- sprzedaży,
- ogólnego zarządu.

Koszty działalności operacyjnej – obejmują koszty ponoszone w celu realizacji podstawowych zadań, dla których przedsiębiorstwo zostało utworzone. W zależności od rodzaju działalności przedsiębiorstwa można wyróżnić:

- koszty działalności produkcyjnej – są to koszty związane z wytworzeniem wyrobów gotowych,
- koszty działalności handlowej – są to koszty związane z zakupem, magazynowaniem i sprzedaż towarów,
- koszty działalności usługowej – są to koszty związane ze świadczeniem usług.

Pozostałe koszty operacyjne – obejmują koszty związane z działalnością operacyjną, ale nie związane z działalnością podstawową. Zaliczyć do nich można:

- wartość netto sprzedanych środków trwałych oraz wartości niematerialnych i prawnych,
- niezawinione niedobory inwentaryzacyjne,
- zapłacone kary, grzywny i odszkodowania,
- przedawnione, umorzone lub nieściągalne należności,
- koszty powstałe przy likwidacji środków trwałych,
- kosztu naprawy szkód komunikacyjnych,

Koszty działalności finansowej – obejmują koszty operacji finansowych, do których zalicza się między innymi prowizje od kredytów, pożyczek, odsetki, wartość zakupu sprzedanych papierów wartościowych, dyskonto przy sprzedaży weksli.

Przykład 2

Przy każdej pozycji kosztu wpisz odpowiedni skrót odpowiadający nazwie rodzaju działalności przedsiębiorstwa: Koszty działalności operacyjnej (KO), pozostała działalność operacyjna (PKO), Działalność finansowa (KF).

Lp.	Informacja	Odpowiedź
1.	Zużyte materiały do produkcji	
2.	Zapłacone odsetki od kredytu	
3.	Koszty naprawy samochodu ciężarowego spowodowane wypadkiem komunikacyjnym	
4.	Ujemne różnice kursowe	
5.	Zapłacone kosztu sądowe związane ze złożonym pozwem sądowym	
6.	Okresowa wymiana oleju i filtrów w samochodzie służbowym	
7.	Zapłacona kara umowna	
8.	Zapłacona faktura za rozmowy telefoniczne	
9.	Zużyte opakowania jednorazowe	
10.	Naprawa maszyny produkcyjnej przez serwis zewnętrzny	

Odpowiedzi

Koszty działalności operacyjnej (KO): 1, 3, 6, 8, 9, 10

Pozostała działalność operacyjna (PKO): 5, 7,

Działalność finansowa (KF): 2, 4,

Koszty działalności podstawowej – obejmują swoim zasięgiem koszty poniesione w związku z wykonywaniem głównej działalności w danej jednostce. Do kosztów tych w szczególności można zaliczyć w zależności od rodzaju działalności:

- produkcyjnej – wartość zużytych materiałów, wynagrodzenia pracowników produkcyjnych, zużycie energii, amortyzacja maszyn;

- handlowej – koszy dzierżawy sklepu, wynagrodzenia pracowników sprzedaży, amortyzacja urządzeń,
- usługowej – koszty zużycia paliw, wynagrodzenia kierowców, amortyzacja ciężarówek.

Koszty działalności pomocniczej – obejmują koszty związane z utrzymaniem komórek uzupełniających, które pośrednio są związane z działalnością gospodarczą danego przedsiębiorstwa. Komórki te wspomagają główna działalność, ale mogą być zastąpione przez jednostki zewnętrzne. W szczególności zaliczamy do nich: wydział remontowy, własna kotłownia, dział transportu, własne ujęcie wody czy oczyszczalnia ścieków, itp.

Koszty sprzedaży – to koszty związane ze sprzedażą wyrobów gotowych, usług czy towarów. W szczególności zaliczymy do nich: koszty reklamy, opakowań jednorazowych, transportu, ubezpieczenia sprzedawanych składników.

Koszty ogólnego zarządu (ogólnozakładowe) – to koszty związane z zarządzaniem i administrowaniem jednostki gospodarczej jako całości przedsiębiorstwa. W szczególności zaliczymy do nich:

- płace wraz z narzutami pracowników zarządu i administracji,
- opłaty pocztowe, bankowe, telekomunikacyjne,
- materiały biurowe, energia elektryczna, cieplna biurowca,
- podróże służbowe zarządu,
- dzierżawa biurowca.

Przykład 3

W przedsiębiorstwie świadczącym usługi przewozu osób koszty bieżącej działalności ujmowane są wyłącznie w przekroju kosztów według miejsc powstawania Koszty działalności podstawowej (DP), Koszty działalności pomocniczej (DPM), Koszty zarządu (KZ). Koszty sprzedaży (KS). Przedsiębiorstwo prowadzi własny warsztat remontowy, w którym dokonuje bieżących napraw i przeglądów autobusów. Informacje dotyczące kosztów bieżącej działalności ujęto w tabeli. Przy każdej pozycji wpisz odpowiedni skrót nazwy kosztu według typu działalności.

Lp.	Informacja	Nazwa kosztu
1.	Wartość zużytego oleju napędowego przez autobusy	
2.	Wynagrodzenia brutto pracowników warsztatu remontowego	
3.	Wynagrodzenia kierowców	
4.	Wynagrodzenia sekretarki	
5.	Zużycie autobusów	
6.	Zużycie komputerów i kserokopiarki w dziale księgowości	
7.	Zużycie budynku warsztatu remontowego	
8.	Koszty odzieży ochronne pracowników warsztatu remontowego	
9.	Zużycie energii elektrycznej w biurowcu zarządu	
10.	Wartość ogłoszenia reklamowego w lokalnej prasie	
11.	Wartość badań technicznych autobusów wykonanych przez stację diagnostyczną	
12.	Zużyci części zamiennych do napraw autobusów	

Odpowiedzi

Koszty działalności podstawowej (DP): 1, 3, 5, 11, 12

Koszty działalności pomocniczej (DPM): 2, 7, 8,

Koszty zarządu (KZ): 4, 6, 9,

Koszty sprzedaży (KS): 10,

3.2. Koszty według rodzajów działalności

Podział kosztów według rodzajów to koszty proste bieżącej działalności podstawowej, które mają charakter jednorodny i nie mogą być już rozłożone na elementarne składniki. W ramach układu rodzajowego wyodrębnia się następujące pozycje:

- amortyzacja,
- zużycie materiałów i energii,
- usługi obce,
- podatki i opłaty,
- wynagrodzenia,
- ubezpieczenia społeczne i inne świadczenia,
- pozostałe koszty.

Zużycie materiałów i energii - zawiera koszty zużycia materiałów podstawowych, paliw, części zamiennych, opakowań, wody, odprowadzania ścieków, gazu, energii elektrycznej i cieplnej. Koszty te obejmują także:

- zużycie materiałów do produkcji, pomocniczych, półproduktów, materiałów biurowych,
- niedobory i ubytki mieszczące się w ustalonych granicach norm.

Usługi obce – to koszty świadczeń mających charakter niematerialny realizowanych przez podmioty zewnętrze. Koszty te obejmują w szczególności usługi:

- transportowe, remontowe, informatyczne, prawnicze, bankowe, pocztowe, telekomunikacyjne, komunalne, podwykonawstwa, leasingu operacyjnego.

Podatki i opłaty – to koszty podatków, które obciążają koszty funkcjonowania przedsiębiorstwa. Koszty te obejmują w szczególności:

- podatek od nieruchomości, rolny, leśny, od środków transportowych,
- podatek akcyzowy,
- podatek od czynności cywilnoprawnych,
- VAT nie podlegający odliczeniu,
- opłaty skarbowe, notarialne.

Wynagrodzenia – to koszty wynagrodzeń za pracę bez względu na stosunek prawny łączący strony. Koszty te obejmują:

- wynagrodzenia osobowe – na podstawie umowy o pracę,
- Wynagrodzenia bezosobowe – na podstawie umów cywilnoprawnych – zlecenie, dzieło, kontrakt menedżerski.

Świadczenia na rzecz pracowników – to koszty tzw. pochodnych wynagrodzeń za pracę oraz związanych z zatrudnieniem pracowników. Koszty te obejmują w szczególności:

- składki ubezpieczeń społecznych, Funduszu Pracy, Funduszu Gwarantowanych Świadczeń Pracowniczych finansowanych przez pracodawcę,
- odpisy na Zakładowy Fundusz Świadczeń Socjalnych,

- pozostałe świadczenia na rzecz pracowników: środki czystości, odzież ochronna, szkolenia pracowników, okresowe badania lekarskie.

Pozostałe koszty rodzajowe – to koszty, które nie zostały zaliczone do w/w grup. Koszty te obejmują w szczególności:

- ubezpieczenia majątkowe,
- Podróże służbowe (krajowe i zagraniczne),
- reklamy,
- koszty reprezentacji

Przykład 4

W przedsiębiorstwie świadczącym usługi precyzyjnej obróbki metalu z wykorzystaniem obrabiarek cyfrowych (wycinanie, toczenie, frezowanie itp.) koszty bieżącej działalności ujmowane są w przekroju kosztów według miejsc powstawania. Koszty działalności podstawowej (DP), Koszty zarządu (KZ), Koszty sprzedaży (KS) oraz według rodzajów: Amortyzacja (A), Zużycie materiałów i energii (ZM), Usługi obce (UO), Podatki i opłaty (P), Wynagrodzenia (W), Świadczenia na rzecz pracowników (SP), Pozostałe koszty rodzajowe (PKR). Informacje dotyczące kosztów bieżącej działalności ujęto w tabeli. Przy każdej pozycji wpisz odpowiedni skrót nazwy kosztu według typu działalności.

Lp.	Informacja	Nazwa kosztu rodzajowego	Koszty według miejsc powstawania
1.	Wartość zużytego oleju do smarowania obrabiarek		
2.	Wartość zużytej energii elektrycznej w dziale produkcyjnym		
3.	Wynagrodzenia brutto pracowników księgowości		
4.	Wynagrodzenia pracowników działu produkcji		
5.	Wynagrodzenie kierownika działu produkcji		
6.	Zużycie obrabiarek		
7.	Naprawa uszkodzonej obrabiarki		
8.	Zużycie budynku produkcyjnego		
9.	Zużyte blachy w dziale produkcyjnym		
10.	Pobrane przez bank prowizje za dokonane przelewy		

11.	Reklama w prasie		
12.	Podatek od nieruchomości budynku produkcyjnego		
13.	Zużycie wody do chłodzenia obrabiarek		
14.	Koszty transportu wykonanych detali do odbiorcy		
15.	Rata leasingu operacyjnego za obrabiarkę		

Odpowiedzi

1 – ZM i DP; 2 – ZM i DP; 3 - W i KZ; 4 - W i DP; 5 - W i DP; 6 – A i DP; 7 – UO i DP; 8 – A i DP; 9 – ZM i DP; 10 – UO i KZ; 11 – PKR i KS; 12 - PO i DP; 13 – ZM o DP; 14 – UO i KS; 15 – UO i DP.

3.3. Pozostałe podziały kosztów

Z punktu widzenia odniesienia kosztów na wyroby gotowe, wyodrębniamy:

- **koszty bezpośrednie –** obejmują koszt, które można, na podstawie dokumentacji źródłowej przypisać do obiektu kosztu (wyrób, usługa), np. kosztu materiałów bezpośrednich, wynagrodzenia bezpośrednie;
- **koszty pośrednie –** obejmują koszty, które nie mogą być odniesione wprost na obiekt kosztów, np. koszty wydziałowe, koszty zarządu, koszty zakupu.

Ważnymi i przydatnymi informacjami dotyczącymi kosztów są te związane z ich zmiennością, która oznacza ich zachowanie spowodowane zmianami określonych parametrów w działalności przedsiębiorstwa. Rozpatrując koszty w relacji do zmian wielkości produkcji, można podzielić ja na:

- **koszty stałe –** obejmują koszty niezależne od zmian wielkości produkcji, pozostające w pewnych granicach na jednakowym poziomie, np. amortyzacja, czynsz, podatek od nieruchomości;
- **koszty zmienne – o**bejmują koszty, które reagują na zmianę w rozmiarach produkcji, np. wynagrodzenia w systemie akordowym, zużycie materiałów bezpośrednich.

Podział kosztów na stałe i zmienne jest bardzo przydatny dla potrzeb decyzyjnych. Jednym ze sposobów podziału kosztów według tego kryterium jest metoda księgowa. Polega ona na tym, że koszty stałe i zmienne wyodrębnia się z kosztów całkowitych w sposób

subiektywny na podstawie doświadczenia, wiedzy i umiejętności osoby dokonującej tego podziału.

Biorą pod uwagę strukturę wewnętrzną kosztów, dzielimy je na:

- **koszty proste –** koszty składające się z jednego elementu (nie da się ich podzielić na mniejsze cząstki) – są to kosztu układu rodzajowego;
- **koszty złożone –** koszt złożony jest sumą kilku kosztów prostych, np. koszt ogólnego zarządu.

W procesie podejmowania decyzji bardzo ważną kwestią jest możliwość kontrolowania kosztów, które możemy podzielić na:

- **koszty kontrolowane –** obejmują koszty, które dają się kontrolować i na które ma wpływ osoba odpowiedzialna za nie;
- **koszty niekontrolowane –** obejmują koszty, które nie da się kontrolować lub można je kontrolować, ale dana osoba nie ma na nie wpływu lub nie ma potrzeby ich kontrolowania.

Z wyżej zaprezentowanym kryterium, związanym z procesem decyzyjnym bardzo zbieżny jest podział kosztów na:

- **koszty decyzyjne –** obejmują koszty, które zmieniają się w wyniku podjęcia danej decyzji;
- **koszty niedecyzyjne –** obejmują koszty, które pozostają niezmienne w wyniku podjęcia danej decyzji.

Koszty możemy rozpatrywać z punktu widzenia regulacji prawnych, z których wynika ich usankcjonowanie. Wówczas można podzielić je na:

- **koszty bilansowe –** obejmują koszty, które zostały określone i zdefiniowane w prawie bilansowym;
- **koszty podatkowe –** obejmują koszty, które zostały określone i zdefiniowane w prawie podatkowym (Prawo podatkowe jest autonomiczne, jak również poszczególne rodzaje podatków posiadają autonomiczną regulację. Powoduje to, że te same pojęcia w poszczególnych podatkach posiadają różniąca się definicję).

Przykład 5

W przedsiębiorstwie wytwarzane są soki owocowe. Pracownicy nakładający nakrętki na butelki pracują w systemie akordowym. W trakcie działalności ponoszone są różne koszty, które ujęto w tabeli. Określ czy dany koszt ma charakter kosztu zmienne go (KZ), czy kosztu stałego (KS).

Lp.	Informacja	Nazwa kosztu
1.	Zużycie wody	
2.	Ubezpieczeni budynku produkcyjnego od zdarzeń losowych	
3.	Podatek od nieruchomości	
4.	Zużycie jabłek, winogron i innych owoców	
5.	Zużycie cukru	
6.	Zużycie komputerów i kserokopiarki w dziale księgowości	
7.	Zużycie budynku produkcyjnego	
8.	Koszty odzieży ochronne pracowników działu produkcyjnego	
9.	Zużycie energii elektrycznej w dziale produkcyjnym	
10.	Wynagrodzenie pracowników zajmujących się nakładaniem nakrętek	

Odpowiedzi:

KS: 2, 3, 6, 7,

KZ: 1, 4, 5, 8, 9, 10.

4. Rozliczenia międzyokresowe kosztów

W procesie agregacji i grupowania kosztów bardzo ważną kwestią jest ich zaliczenie do właściwego okresu. Zdarza się, że poniesione koszty obejmują więcej niż jeden okres rozliczeniowy lub istnieje potrzeba zarachowania ich wcześniej, wiedząc, że ich poniesienie jest wysoce prawdopodobne. W takich sytuacjach konieczne jest stosowanie rozliczeń międzyokresowych kosztów. Rozliczenia międzyokresowe kosztów przeprowadza się w celu ścisłego ustalenia kosztów działalności dla każdego okresu sprawozdawczego bez względu na to, czy odniesione koszty zostały poniesione w poprzednich okresach sprawozdawczych, czy też odpowiadające bieżącym kosztom wydatki będą poniesione dopiero w przyszłych okresach.

Rozliczenia te dzielimy na:

- **czynne rozliczenie międzyokresowe kosztów** – obejmują koszty, które zostały poniesione w jednym okresie rozliczeniowym (np. miesiącu, roku), a dotyczą następnych okresów rozliczeniowych;
- **bierne rozliczenie międzyokresowe kosztów** – obejmują koszty, które dotyczą danego okresu rozliczeniowego, a zostaną poniesione w następnym okresie rozliczeniowym.

Stosowanie rozliczeń międzyokresowych kosztów wynika z konieczności przestrzegania kompletności kosztów, tj. zaliczenia do kosztów własnych kosztów związanych z danym okresem oraz zasada ich prawidłowego rozliczenia w czasie, czyli ścisłego rozgraniczenia kosztów okresu sprawozdawczego.

5. Rozliczenia kosztów pośrednich

Koszty pośrednie w przeciwieństwie do kosztów bezpośrednich nie dają możliwości automatycznego przyporządkowania ich do poszczególnych obiektów odniesienia kosztów. Koszty pośrednie, takie jak:

- koszty zakupu,
- koszty wydziałowe,
- koszty zarządu,
- koszty sprzedaży

rozlicza się na wyroby, usługi między innymi za pomocą kluczy rozliczeniowych (kluczy podziałowych).

Klucz rozliczeniowy kosztów pośrednich jest to umowna wielkość, stanowiąca podstawę rozliczenia kosztów pośrednich na poszczególne obiekty odniesienia kosztów.
Niezmiernie ważne jest, aby dla danych kosztów pośrednich dobrać odpowiedni klucz rozliczeniowy.

Aby dana wielkość mogła pełnić rolę klucza rozliczeniowego kosztów pośrednich muszą odpowiadać następującym warunkom:

- klucz rozliczeniowy powinien pozostać w związku przyczynowo-skutkowym z rozliczanymi kosztami,
- klucz rozliczeniowy powinien pozostać w związku proporcjonalnym z rozliczanymi kosztami,

- związek klucza rozliczeniowego z kosztami podlegającymi rozliczeniu powinien być możliwie silny,
- klucz rozliczeniowy powinien być parametrem charakteryzującym te obiekty odniesienia kosztów, na które koszty pośrednie będą rozliczane;
- musi istnieć możliwość jednoznacznego przyporządkowania liczby jednostek klucza obiektom odniesienia kosztów.

Klucze rozliczeniowe kosztów można klasyfikować według różnych kryteriów. Z punktu widzenia stabilności wielkości klucza rozliczeniowego, są one dzielona na:

- stałe – to takie których ogólna suma jednostek jest niezmienna w ciągu dłuższego okresu. Mogą one wynikać z planu techniczno-ekonomicznego jednostki gospodarczej.
- zmienne – charakteryzujące się tym, że w każdym okresie zmienia się zarówno ogólna suma jednostek klucza, jak ich liczba przypadająca na poszczególne obiekty odniesienia kosztów.

Biorąc pod uwagę jednostkę miary, wyrażającą wielkość klucza rozliczeniowego, można podzielić ja na:

- wartościowe – wyrażone są w jednostkach pieniężnych, np. robocizna bezpośrednia wyrażona w złotówkach,
- ilościowe – wyrażone w naturalnych jednostkach miary, np. liczba roboczogodzin, maszynogodzin. (P. Szczypa, 2017)

Dobór kluczy podziałowych kosztów jest uzależniony od różnych czynników. W dużej mierze jest to rodzaj prowadzonej działalności, rodzaj wytwarzanych produktów, czy świadczonych usług. (G. Lew, E. Maruszewska, P. Szczypa, 2025).

6. Kalkulacja kosztów

Kalkulacja kosztów stanowi istotny element rachunku kosztów. Wynik kalkulacji zależy nie tylko od zastosowanej metody, ale przede wszystkim od przyjętych zasad grupowania, klasyfikacji i alokacji kosztów, co z kolei jest uzależnione od modelu rachunku kosztów (G. Dykto 2000).

Poniesione przez jednostkę koszty okresu sprawozdawczego podlegają podziałowi na produkcje gotową i produkcje niezakończoną, a także na poszczególne produkty. **Kalkulacja** jest czynnością obliczeniową, zmierzającą do ustalenia kwoty kosztów przypadających na przedmiot kalkulacji. (W. Gabrusewicz, A. Kamela Sowińska, H. Poetschke, 1996).

Produkty gotowe, półprodukty i produkcji w toku mogą być wyceniane według:

- kosztów zmiennych,
- technicznego kosztu wytworzenia: kosztów bezpośrednich + kosztów wydziałowych,
- kosztów pełnych: technicznego kosztu wytworzenia (koszty bezpośrednie + koszty wydziałowe) + kosztów ogólnozakładowych (koszty zarządu).

Kalkulacja jest elementem rachunku kosztów, której zadaniem jest podział kosztów według pozycji kalkulacyjnych na przedmiot kalkulacji. Przedmiotem kalkulacji (jednostka kalkulacyjna, obiekt kosztów) może być między innymi:

- jednostka wyrobu gotowego (np. kilogram wyprodukowanego masła),
- jednostka usługi (np. wozokilometr w transporcie towarów),
- jednostka zlecenia produkcyjnego (np. zlecenie wyprodukowania dodatkowej partii 100 szt. spodni).

Kalkulacje dzielimy na:

- wstępną – jest ona sporządzana przed rozpoczęciem produkcji wyrobu i opiera się na kosztach przewidywanych,
- wynikową – jest sporządzana po zakończeniu produkcji i opiera się na kosztach rzeczywistych zarejestrowanych na kontach księgowych.

Kalkulacja wstępna jest na wielkościach przewidywanych, planowanych (przewidywane wielkości produkcji, świadczenia usług, normy zużycia materiałów, przewidywane ceny materiałów, usług obcych, wynagrodzeń, itd.). W ramach kalkulacji wstępnej można wyodrębnić:

- kalkulację planowaną – opartą na wielkościach planowanych,
- kalkulację normatywną – opartą na ustalonych normach zużycia materiałów, normach czasu pracy),

- kalkulacje ofertową – oparta na wielkościach szacunkowych, zgodnie z poziomem cen w danym regionie.

Kalkulacja wynikowa (sprawozdawcza, rzeczywista) oparta jest na rzeczywistych kosztach, które zostały poniesione w danym okresie. Dzięki porównaniu kosztów jednostkowych ustalonych w kalkulacji wstępnej i wynikowej można ustalić różnice między kosztami rzeczywistymi a kosztami przewidywanymi. Wyciągnięcie odpowiednich wniosków z takiego porównania może przyczynić się do obniżenia kosztów prowadzonej działalności. Stanowi ono podstawę kontroli wykonania planu kosztów w zakresie poszczególnych wyrobów. Przyjmując za kryterium podziału kalkulacji kompletność kosztów uwzględnionych przy ustalaniu kosztu jednostkowego, można wyodrębnić:

- kalkulację kosztu częściowego – kalkulacja ta znajduje zastosowanie w rachunku kosztów zmiennych, której podstawę stanowią jedynie koszty zmienne,
- kalkulację kosztu pełnego – kalkulacja ta znajduje zastosowanie w rachunku kosztów pełnych, której podstawę stanowią wszystkie poniesione na przedmiot kalkulacji koszty, zarówno bezpośrednie jak i pośrednie.

Koszt jednostkowy może być ustalony za pomocą różnych metod. Dobór odpowiedniej metody jest uzależniony od:

- przedmiotu kalkulacji,
- przebiegu i organizacji procesu technologicznego,
- charakteru wytwarzanego produktu.

Biorąc po uwagę metody odnoszenia kosztów na jednostkę produkcji, kalkulację dzielimy na:

- podziałowa:
  - prosta,
  - współczynnikowa,
  - procesowa – półfabrykatowa i bezpółfabrykatowa,
- doliczeniowa:
  - zleceniowa,
  - asortymentowa.

6.1. Kalkulacja podziałowa

Kalkulacja podziałowa jest najprostszą i najstarszą metodą kalkulacji. Jest stosowana w wypadku produkcji prostej, gdy wyroby są wytwarzane na ogół w jednym, nieprzerwanym cyklu produkcyjnym. Produkcja prosta dotyczy jedynie produkcji masowe, np. produkcji drutu, gwoździ, opakowań, piwa, cukru itp.

Kalkulacja podziałowa prosta jest stosowana w przedsiębiorstwach wytwarzających masowo jednorodny produkt (wyrób gotowy lub usługę), np. kopalnie węgla, elektrownie, gorzelnie, cukrownie itp.

Przykład 6

Zgrupowane są następujące wartości kosztów:

- materiały bezpośrednie – 1000 zł,
- płace bezpośrednie – 700 zł,
- koszty wydziałowe – 500 zł

W bieżącym miesiącu wytworzono 80 szt. jednorodnych wyrobów gotowych oraz 40 szt. Półfabrykatów, które z punktu widzenia poniesionych kosztów produkcji przerobiono w 50% w stosunku do wyrobu gotowego. Ustal koszt jednostkowy wyrobu gotowego i półfabrykat.

Rozwiązanie

Ustalenie ilości jednostek kalkulacyjnych: 80 + (40 x 50%) = 80 + 20 = 100 szt.

Ustalenie łącznych kosztów wytworzenia wyrobów gotowych: 1000 + 700 + 500 + 2200 zł

Ustalenie kosztu jednostkowego wyrobu gotowego: 2200 : 100 = 220 zł

Ustalenie kosztu jednostkowego półfabrykatu: 220 x 50% = 110 zł

Kalkulacja podziałowa współczynnikowa jest stosowana w przedsiębiorstwach o produkcji masowej wieloasortymentowej. Przy założeniu, że produkcja ta wymaga dla wszystkich wytwarzanych rodzajów wyrobów:

- takich samych lub podobnym materiałów,
- podobnym urządzeń produkcyjnych,
- takich samych procesów technologicznych.

Może mieć zastosowanie w: piekarni, hucie szkła, wytwórni wyrobów z mas plastycznych, wytwórni okien, produkcji makaronu itp.

Istota kalkulacji podziałowej współczynnikowej polega na sprowadzeniu różnych wyrobów do wspólnego mianownika poprzez przeliczenie ich na umowne obiekty kalkulacyjne za pomocą współczynników.

Przykład 7

Zakład przemysłu metalowego produkuje z blachy dwa rodzaje pudełek do pasty o pojemności 200 g i 300 g. W ciągu bieżącego okresu wyprodukowano 10200 pudełek o pojemności 200 g oraz 4000 pudełek o pojemności 300 g. za przedmiot kalkulacji przyjmuje się 100 pudełek każdego rodzaju. Koszty produkcji są proporcjonalne do wielości pudełek. Poniesione w ciągu miesiąca koszty produkcji pudełek są następujące:

- materiały bezpośrednie – 8100 zł
- płace bezpośrednie – 3240 zł
- koszty wydziałowe – 2268 zł

Ustal koszt jednostkowy jednostki kalkulacyjnej.

Rozwiązanie

Wyrób	**Liczba wytworzona**	**Współczynnik**	**Liczba jednostek współczynnikowych (kol 2x3)**	**Koszt jednostki współczynnikowej w zł**	**Koszt wyrobu gotowego (kol. 3x5) w zł**	**Koszt wykonanej produkcji (kol. 2x6) w zł**
1	2	3	4	5	6	7
Materiały bezpośrednie:						
-pudełka 200g	102	2	204	25	50	5100
-pudełka 300g	40	3	120	25	75	3000
			Razem: 324			
Materiały bezpośrednie:						
-pudełka 200g	102	2	204	10	20	2040
-pudełka 300g	40	3	120	10	30	1200
			Razem: 324			

Materiały bezpośrednie:						
-pudełka 200g	102	2	204	7	14	1428
-pudełka 300g	40	3	120	7	21	840
			Razem: 324			

Dodatkowe obliczenia:

1. Obliczenie kosztów przypadających na jednostkę współczynnikową:
   - materiały bezpośrednie 8100 : 324 = 25 zł
   - płace bezpośrednie 3240 : 324 = 10 zł
   - koszty wydziałowe 2268 : 324 = 7 zł
2. Ustalenie łącznego kosztu wytworzenia wyrobów gotowych:

	Pudełka 200 g	Pudełka 300g
– materiały bezpośrednie	50 zł	75 zł
– płace bezpośrednie	20 zł	30 zł
– koszty wydziałowe	14 zł	21 zł
– koszty wytworzenia wyrobu	84 zł	126 zł

Kalkulacja podziałowa procesowa (kalkulacja podziałowa fazowa, kalkulacja podziałowa wielostopniowa) powinna być stosowana w przedsiębiorstwach o produkcji masowej i wieloseryjnej, w których proces produkcji obejmuje kilka następujących po sobie faz (etapów) produkcji. W każdej fazie produkcji powstaje półfabrykat, który jest przekazywany do dalszej fazy produkcji i tak aż do ostatniej fazy, gdzie powstaje wyrób gotowy. Metoda ta jest stosowana w cegielniach, przy produkcji samochodów.

W trakcie sporządzania kalkulacji podziałowej procesowej wykorzystuje się procedury kalkulacji podziałowej prostej lub ze współczynnikami. Można ją przeprowadzić w dwóch wariantach:

- jako kalkulację bezproduktową, polegającą na tym, że w każdej fazie ustala się oddzielnie jednostkowy koszt wytworzenia półproduktu, a następnie sumuje się poszczególne jednostkowe koszty z danych faz produkcji otrzymując w ten sposób jednostkowy koszt wyrobu gotowego,
- jako kalkulację półproduktową, która polega na tym, że koszty wytworzenia półproduktów z danej fazy dolicza się do kolejnej. Taki proces zapewnia ustalenie kosztu jednostkowego w sposób narastający.

6.2. Kalkulacja doliczeniowa

Kalkulacja doliczeniowa jest stosowana w produkcji złożonej, montażowej, która charakteryzuje się tym, że produkowane wyroby składają się z wielu części produkowanych i montowanych w jednym lub kilku wydziałach. W zależności od organizacji wydziałów produkcyjnych mogą one być tworzone według struktury przedmiotowej lub technologicznej. Za pomocą kalkulacji doliczeniowej ustala się więc na ogół indywidualne bezpośrednie koszty wytworzenia danego wyrobu (roboty lub usługi), z tym, że w produkcji seryjnej są to indywidualne koszty bezpośrednie danej serii wyrobu. Dla każdego wyrobu objętego serią otrzymuje się, jak w kalkulacji podziałowej, przeciętny koszt wytworzenia.

W zależności od rodzaju produkcji oraz jej organizacji mogą występować dwie odmiany kalkulacji doliczeniowej:

- **kalkulacja doliczeniowa zleceniowa** – stosowana w produkcji jednostkowej małoseryjnej. Przedmiotem kalkulacji jest konkretne zlecenie produkcyjne, obejmujące część składową wyrobu gotowego, wyrób gotowy lub serię wyrobów gotowych;
- **kalkulacja doliczeniowa asortymentowa** – stosowana w produkcji seryjnej wieloseryjnej złożonej (montażowej). Przedmiotem kalkulacji jest konkretny asortyment wyrobów gotowych, dla którego otwiera się odpowiednie urządzenie ewidencyjne, służące do ujmowania kosztów związanych z produkcją danego asortymentu.

Przykład 8

W miesiącu sprawozdawczym jednostka gospodarcza wykonywała produkcję na dwa zlecenia produkcyjne:

- Zlecenie nr 1 – 5 szt. wyrobu X
- Zlecenie nr 2 – 10 szt. wyrobu Y

Koszty produkcji zaewidencjonowane na odpowiednim koncie księgowym w przekroju pozycji kalkulacyjnych przedstawiają się następująco:

Koszty bezpośrednie	Ogółem	Zlecenie nr 1	Zlecenie nr 2
Materiały bezpośrednie (zł)	20 000	8 000	12 000
Płace bezpośrednie (zł)	8 000	3 000	5 000
Koszty wydziałowe (zł)	12 000		
Razem	40 000		

Na podstawie przedstawionych danych należy sporządzić kalkulacje wyrobów gotowych wiedząc, że wszystkie wyroby zostały zakończone, a koszty wydziałowe są rozliczne w stosunku do płac bezpośrednich.

Na podstawie podanych informacji:

1. Ustal koszt wytworzenia zlecenia.
2. Ustal koszt jednostkowy wyrobu X i Y.

Rozwiązanie

Wskaźnik narzutu kosztów pośrednich:

$$Wn = \frac{12000\ x100}{8000} = 150\%$$

Koszty wydziałowe przypadające na:

- Zlecenie 1: 150% x 3 000 = 4 500
- Zlecenie 2: 150% x 5 000 = 7 500

Koszty bezpośrednie	Zlecenie nr 1	Zlecenie nr 2
Materiały bezpośrednie (zł)	8 000	12 000
Płace bezpośrednie (zł)	3 000	5 000
Koszty wydziałowe (zł)	4 500	7 500
Razem koszt wytworzenia	15 500	24 500
Liczba wytworzonych wyrobów gotowych	5	10
Jednostkowy koszt wytworzenia	3 100	2 450

7. Rachunek kosztów

W dążeniu do coraz lepszego spełnienia zadań, jakie stawiane są rachunkowi kosztów, wykształciły się następujące rodzaje rachunku kosztów, stosowane najczęściej:

- rachunek kosztów pełnych,
- rachunek kosztów zmiennych,
- rachunek kosztów standardowych,
- rachunek kosztów działań.

Rachunek kosztów pełnych jest oparty na kosztach historycznych, czyli już poniesionych. Jest on też często nazywany rachunkiem tradycyjnym. Zakłada, że na

wysokość całkowitych kosztów przedsiębiorstwa wpływ ma jedna zmienna – wielkość produkcji. Pomija on jednak to, że w gospodarce rynkowej stopień wykorzystania zdolności produkcyjnych przedsiębiorstwa może być różny, a niekiedy bardzo niski. Wówczas niekorzystnie mogą się układać proporcje pomiędzy kosztami stałymi i zmiennymi.

Do najważniejszych czynników powodujących wzrost kosztów pośrednich należą:

- zastąpienie pracy żywej pracą maszyn,
- zmiana charakteru kosztów płac z proporcjonalnie zmiennych na koszty stałe,
- wzrost złożoności prac faz przedprodukcyjnych, związanych z np. wprowadzeniem nowego wyrobu,
- wzrost kosztów fazy poprodukcyjnej, w tym głównie sprzedaży,

Powszechność stosowania rachunku kosztów pełnych wynika z wymogów prawa bilansowego, a w przypadku rachunku kosztów zmiennych ze względu na jego nieskomplikowaną strukturę i dużą przydatność przy podejmowaniu decyzji (P. Szczypa, 2017).

Istotą **rachunku kosztów zmiennych** jest podział całkowitych kosztów własnych przedsiębiorstwa na koszty zmienne, proporcjonalne do wielkości produkcji oraz na koszty stałe, związane z określonym przedziałem czasu. W konsekwencji takiego podziału kosztów produktom przypisuje się tylko koszty zmienne, w przeciwieństwie do rachunku kosztów pełnych. Koszty stałe w rachunku kosztów zmiennych nie obciążają produktów, lecz są traktowane jako koszty okresu sprawozdawczego i w całości odnoszone do wyniku tego okresu.

$$KC=\sum KZ_j \text{ x } Q + KS$$

Tabela 1. Procedura ustalania wyniku ze sprzedaży w rachunku kosztów zmiennych przy produkcji jednoasortymentowej

Lp.	Treść	Wyrób A
1.	Przychody ze sprzedaży wyrobów gotowych (usług, towarów)	X
2.	Koszty zmienne wyrobów gotowych (usług, towarów)	X
3.	Marża brutto (pozycja 1 – pozycja 2)	X
4.	Koszty stałe	X
5.	Wynik ze sprzedaży (pozycja 3 – pozycja 4)	X

Źródło: P. Szczypa, 2017.

Tabela 2. Procedura ustalania wyniku ze sprzedaży w rachunku kosztów zmiennych przy produkcji wieloasortymentowej

Lp.	Treść	Wyrób A	Wyrób B	Wyrób C
1.	Przychody ze sprzedaży wyrobów gotowych (usług, towarów)	X	X	X
2.	Koszty zmienne wyrobów gotowych (usług, towarów)	X	X	X
3.	Marża brutto (pozycja 1 – pozycja 2)	X	X	X
4.	Koszty stałe	X		
5.	Wynik ze sprzedaży (pozycja 3 – pozycja 4)	X		

Źródło: P. Szczypa, 2017.

Rachunek kosztów standardowych jest rachunkiem kosztów apriorycznych, tj. opartym na wartościach kosztów ustalonych z góry, wartościach planowanych. Zgodnie z tym modelem koszty rzeczywiste (Kr) stanowią wartość kosztów apriorycznych (Ka) skorygowanych o odchylenia:

$$Kr = Ka+/- O$$

Można wyróżnić dwie odmiany rachunku kosztów standardowych:

- rachunek kosztów normatywnych
- rachunek kosztów normalnych

Rachunek kosztów normatywnych wykorzystuje ściśle określone technicznie normy zużycia bezpośrednich i pośrednich czynników wytwórczych. **W rachunku kosztów normalnych** poszczególne wartości ustalane są za pomocą procedur statystycznych np. jako średnia wartość danych kosztów z kilku poprzednich okresów sprawozdawczych. Koncentrowanie się kadry zarządzającej na procesach i działaniach zachodzących w przedsiębiorstwie znalazło swoje przełożenie na rachunek kosztów, określanego mianem **rachunku kosztów działań** lub koncepcją ABC (ang. *Activity Based Costing*). Zgodnie z tą koncepcją, koszty pośrednie rozlicza się na produkty w przekroju działań i procesów generujących te koszty, a nie w przekroju podmiotów produkcyjnych (wydziałów), przy zastosowaniu wielu różnych podstaw rozliczenia, które w większości nie są proporcjonalne do wielkości produkcji. Fundamentalnym założeniem tej koncepcji jest to, że działania powodują powstawanie kosztów, a wyroby lub usługi kreują zapotrzebowanie na te działania. Do najważniejszych motywów skłaniających przedsiębiorstwa do wdrożenia rachunku kosztów działań jest zidentyfikowanie tych działań, które nie przynoszą wartości firmy.

Podsumowanie

W rozdziale zaprezentowano podstawowe zagadnienia z zakresu rachunku kosztów, przedstawiając je w sposób skierowany dla odbiorców niebędących ekonomistami. Przedstawiono podział kosztów według różnych kryteriów, sposoby ich grupowania, a także rozliczania z wykorzystaniem kalkulacji w celu ustalania kosztów wytworzenia wyrobów gotowych lub świadczonych usług. Syntetycznie scharakteryzowano różne modele rachunku kosztów, ze wskazaniem ich podstawowych cech oraz roli w procesie decyzyjnym. Zawarte w rozdziale informacje mają na celu przybliżyć i ułatwić zrozumienie zagadnień związanych z rachunkiem kosztów, w szczególności przez kadrę menedżerską działów technicznych i produkcyjnych oraz właścicieli mikro i małych przedsiębiorstw.

Bibliografia

Gabrusewicz, W., Kamela-Sowińska, A., Poetschke, H. (1996). Rachunkowość zarządcza, Poznań: Wydawnictwo Akademii Ekonomicznej w Poznaniu.

Jaruga, A., Nowak, W., Szychta, A. (2001), Rachunkowość zarządcza koncepcje i zastosowania, Łódź: Społeczna Wyższa Szkoła Przedsiębiorczości i Zarządzania w Łodzi

Karmańska, A. (red.) (2022), Rachunkowość zarządcza i rachunek kosztów w systemie informacyjnym przedsiębiorstwa, Warszawa: SGH Oficyna Wydawnicza.

Kołaczyk, Z. (1996), Rachunkowość finansowa, Poznań: Wydawnictwo Akademii Ekonomicznej w Poznaniu.

Lew, G., Maruszewska, E., Szczypa, P. (2025), Rachunkowość zarządcza od teorii do praktyki, Warszawa: CeDeWu.

Matuszewicz, J. (2002), Rachunek kosztów, Warszawa: Finans-Servis.

Nowak E. (2018), Rachunkowość zarządcza w przedsiębiorstwie, Warszawa: CeDeWu.

Sojak, S., Jóźwiak, H. (2004), Rachunek kosztów docelowych, Kraków: Oficyna Wydawnicza.

Szczypa, P. (2017), Rachunkowość zarządcza – klucz do sukcesu, Warszawa: CeDeWu.

Walińska, E., Urbanek, P. (2000), Rachunkowość zarządcza wybrane zagadnienia testy i zadania, Łódź: Fundacja Rozwoju rachunkowości w Polsce.

Fundusze Europejskie dla Wielkopolski
Dofinansowane przez Unię Europejską
SAMORZĄD WOJEWÓDZTWA WIELKOPOLSKIEGO

Innowacje i usprawnienia w firmach

dr inż. Monika Muszyńska

Wstęp

Współczesne przedsiębiorstwa funkcjonują w warunkach dynamicznych i często nieprzewidywalnych zmian otoczenia gospodarczego, technologicznego oraz społecznego. Globalizacja rynków, rosnąca konkurencja, rozwój nowych technologii oraz zmieniające się potrzeby klientów sprawiają, że organizacje są zmuszone do ciągłego doskonalenia swojej działalności. W tym kontekście szczególnego znaczenia nabierają innowacje oraz usprawnienia procesów, które stanowią fundament budowania przewagi konkurencyjnej oraz długoterminowego rozwoju przedsiębiorstw (Wiśniewska, Janasz 2023).

Innowacyjność przedsiębiorstw jest obecnie jednym z kluczowych czynników determinujących ich zdolność do przetrwania i rozwoju w warunkach gospodarki opartej na wiedzy. Jak wskazują liczne opracowania, innowacje nie ograniczają się jedynie do tworzenia nowych produktów czy technologii, lecz obejmują również zmiany w procesach produkcyjnych, organizacji pracy, modelach biznesowych czy działaniach marketingowych (Trzepizur 2016). Tak szerokie ujęcie innowacyjności pozwala na kompleksowe spojrzenie na funkcjonowanie przedsiębiorstwa jako systemu, w którym każdy element może być przedmiotem usprawnienia.

Zgodnie z klasycznym ujęciem zaproponowanym przez Schumpetera, innowacje obejmują m.in. wprowadzanie nowych produktów, udoskonalanie procesów produkcyjnych, otwieranie nowych rynków czy wdrażanie nowych form organizacji działalności gospodarczej (Schumpeter 1960). Współczesne podejścia rozwijają tę koncepcję, wskazując na cztery podstawowe typy innowacji: produktowe, procesowe, organizacyjne oraz marketingowe (OECD 2008). Każdy z tych typów odgrywa istotną rolę w kształtowaniu efektywności przedsiębiorstwa oraz jego zdolności adaptacyjnych.

Znaczenie innowacji we współczesnej gospodarce podkreślane jest również w kontekście rozwoju społeczno-gospodarczego państw i regionów. Innowacyjność stanowi bowiem podstawę konkurencyjności gospodarek oraz kluczowy element strategii rozwoju na poziomie krajowym i regionalnym (Matusiak 2006). W warunkach integracji europejskiej oraz globalnych powiązań gospodarczych, zdolność do generowania i wdrażania innowacji

decyduje o pozycji konkurencyjnej zarówno pojedynczych przedsiębiorstw, jak i całych gospodarek. W literaturze przedmiotu podkreśla się, że innowacje są nierozerwalnie związane z kreatywnością oraz zdolnością organizacji do uczenia się i adaptacji. Kreatywność stanowi źródło nowych pomysłów, które w procesie ich rozwijania i wdrażania przekształcają się w innowacje przynoszące wymierne efekty ekonomiczne (Wiśniewska, Janasz 2023). Tym samym przedsiębiorstwa, które potrafią skutecznie zarządzać wiedzą oraz potencjałem twórczym pracowników, osiągają wyższy poziom innowacyjności.

W kontekście niniejszego opracowania szczególne znaczenie mają również usprawnienia procesów realizowanych w przedsiębiorstwie. Usprawnienia te, często określane jako innowacje inkrementalne, polegają na systematycznym doskonaleniu istniejących rozwiązań, co prowadzi do zwiększenia efektywności, redukcji kosztów oraz poprawy jakości produktów i usług. W praktyce gospodarczej działania te są realizowane m.in. poprzez analizę procesów, identyfikację problemów oraz wdrażanie narzędzi doskonalenia, takich jak analiza strumienia wartości czy mapowanie procesów.

Zgodnie z założeniami przedmiotu „Innowacje i usprawnienia w firmach", kluczowym obszarem kształcenia jest nabycie przez studentów umiejętności analizy funkcjonowania przedsiębiorstwa oraz identyfikacji obszarów wymagających usprawnień. Szczególny nacisk kładziony jest na analizę techniczno-ekonomiczną procesów oraz wykorzystanie narzędzi pozwalających na ocenę ich efektywności. Takie podejście umożliwia studentom nie tylko zrozumienie istoty innowacji, ale również praktyczne zastosowanie zdobytej wiedzy w realnych warunkach przedsiębiorstwa.

Współczesne przedsiębiorstwa, zwłaszcza działające w sektorze przemysłowym, coraz częściej wykorzystują podejście procesowe do zarządzania. Oznacza to konieczność identyfikacji, analizy oraz optymalizacji procesów podstawowych i pomocniczych. W tym kontekście szczególnie istotne staje się wykorzystanie narzędzi takich jak mapa strumienia wartości (*Value Stream Mapping),* która pozwala na identyfikację działań generujących wartość oraz eliminację marnotrawstwa. Umiejętność stosowania tego typu narzędzi stanowi jeden z kluczowych efektów kształcenia w ramach analizowanego przedmiotu.

Nie bez znaczenia pozostaje również rola pracowników przedsiębiorstwa w procesie doskonalenia organizacji. Współczesne koncepcje zarządzania podkreślają znaczenie zaangażowania pracowników w procesy innowacyjne oraz ich udział w identyfikacji problemów i proponowaniu rozwiązań. Budowanie kultury organizacyjnej sprzyjającej

innowacyjności oraz ciągłemu doskonaleniu stanowi jeden z kluczowych elementów sukcesu przedsiębiorstw (Kubiński 2017).

Warto również zwrócić uwagę na specyfikę innowacji w małych i średnich przedsiębiorstwach, które często dysponują ograniczonymi zasobami finansowymi oraz kadrowymi. Pomimo tych ograniczeń, sektor MŚP odgrywa istotną rolę w gospodarce, a jego rozwój jest silnie uzależniony od zdolności do wdrażania innowacji (Trzepizur 2016). W związku z tym szczególnego znaczenia nabiera efektywne zarządzanie procesem innowacyjnym, obejmujące etapy od generowania pomysłów po ich wdrożenie i kontrolę efektów. Istotnym aspektem omawianej problematyki jest także konieczność ciągłego doskonalenia oraz uczenia się przez całe życie. W dynamicznie zmieniającym się otoczeniu gospodarczym przedsiębiorstwa muszą nieustannie podnosić swoje kompetencje oraz dostosowywać się do nowych warunków. W tym kontekście rozwijanie postawy proinnowacyjnej oraz umiejętności analitycznych wśród studentów stanowi jeden z głównych celów kształcenia.

Podsumowując, innowacje i usprawnienia w przedsiębiorstwach stanowią kluczowy obszar badań oraz praktyki gospodarczej, determinujący zdolność organizacji do konkurowania na rynku oraz adaptacji do zmieniających się warunków otoczenia. Kompleksowe podejście do tej problematyki, obejmujące zarówno aspekty teoretyczne, jak i praktyczne, umożliwia lepsze zrozumienie mechanizmów funkcjonowania współczesnych przedsiębiorstw oraz przygotowanie do podejmowania decyzji w zakresie ich doskonalenia.

Niniejsze opracowanie ma na celu przedstawienie istoty innowacji i usprawnień w przedsiębiorstwach, ze szczególnym uwzględnieniem metod analizy procesów oraz narzędzi ich doskonalenia. W kolejnych rozdziałach zaprezentowane zostaną wybrane koncepcje teoretyczne oraz przykłady praktycznych zastosowań, które pozwolą na pogłębienie wiedzy oraz rozwinięcie umiejętności niezbędnych w pracy inżyniera i menedżera.

1. Istota, znaczenie i rodzaje innowacji oraz usprawnień w przedsiębiorstwie

1.1. Wprowadzenie

Innowacje i usprawnienia należą dziś do podstawowych kategorii opisujących rozwój współczesnych przedsiębiorstw. Organizacje funkcjonujące w warunkach rosnącej konkurencji, presji kosztowej, szybkiego postępu technologicznego oraz zmiennych oczekiwań

klientów nie mogą ograniczać się jedynie do utrzymywania dotychczasowych sposobów działania. Aby zachować zdolność konkurowania, muszą stale poszukiwać nowych rozwiązań, udoskonalać procesy, modernizować produkty i doskonalić organizację pracy. Taka perspektywa jest zgodna z założeniami przedmiotu „Innowacje i usprawnienia w firmach", w którym podkreśla się zarówno znaczenie innowacji we współczesnym przedsiębiorstwie, jak i rolę pracowników w ciągłym doskonaleniu procesów (Wiśniewska, Janasz 2023).

W praktyce gospodarczej pojęcia „innowacja" i „usprawnienie" są często używane zamiennie, choć nie oznaczają dokładnie tego samego. Innowacja zwykle odnosi się do wdrożenia rozwiązania nowego lub znacząco udoskonalonego, natomiast usprawnienie częściej wiąże się z poprawą istniejącego sposobu działania, zwiększeniem efektywności, skróceniem czasu realizacji procesu, poprawą jakości albo ograniczeniem kosztów. W rzeczywistości oba zjawiska są ze sobą silnie powiązane. Wiele dużych zmian w przedsiębiorstwie zaczyna się od szeregu drobnych usprawnień, a liczne innowacje mają charakter przyrostowy, a nie przełomowy. Taki sposób rozumienia zmian jest szczególnie ważny dla studentów kierunków technicznych, ponieważ pozwala spojrzeć na przedsiębiorstwo nie tylko jako na miejsce wdrażania spektakularnych technologii, ale również jako na przestrzeń systematycznej analizy i doskonalenia codziennych działań.

1.2. Pojęcie „innowacji" w literaturze przedmiotu

Pojęcie innowacji jest szeroko analizowane w literaturze ekonomicznej i zarządczej. Najczęściej rozumiane jest jako wprowadzenie nowego lub istotnie ulepszonego rozwiązania w obszarze produktu, procesu, organizacji lub rynku. W ujęciu klasycznym innowacja obejmuje zarówno sam rezultat (np. nowy produkt), jak i proces jego powstawania oraz wdrożenia (Trzepizur 2016). Jedną z najważniejszych koncepcji innowacji zaproponował Schumpeter, który wskazał, że innowacje obejmują m.in. wprowadzanie nowych produktów, doskonalenie metod produkcji, otwieranie nowych rynków czy wdrażanie nowych form organizacyjnych (Schumpeter 1960). Podejście to jest nadal aktualne i stanowi punkt odniesienia dla współczesnych analiz innowacyjności. Współczesne ujęcia podkreślają, że innowacja nie musi mieć charakteru przełomowego, może być nowa jedynie z perspektywy danego przedsiębiorstwa, ale nadal przynosić istotne efekty ekonomiczne (Kubiński 2017). Co istotne, wdrażanie innowacji zawsze wiąże się z ryzykiem oraz koniecznością podejmowania decyzji w warunkach niepewności.

1.3. Innowacyjność jako zdolność przedsiębiorstwa

Z pojęciem innowacji bezpośrednio związana jest innowacyjność, rozumiana jako zdolność przedsiębiorstwa do tworzenia, wdrażania i wykorzystywania nowych rozwiązań. Innowacyjność nie jest jednorazowym działaniem, lecz trwałą cechą organizacji, wynikającą z jej kultury organizacyjnej, kompetencji pracowników oraz sposobu zarządzania (Wiśniewska, Janasz 2023).

W literaturze podkreśla się, że innowacyjność stanowi jeden z najważniejszych czynników rozwoju przedsiębiorstw oraz gospodarek. Jest ona podstawą konkurencyjności oraz warunkiem skutecznego funkcjonowania w warunkach gospodarki opartej na wiedzy (Matusiak 2006). Przedsiębiorstwa zdolne do wdrażania innowacji szybciej adaptują się do zmian oraz lepiej wykorzystują pojawiające się szanse rynkowe.

Z perspektywy dydaktycznej istotne jest, aby student rozumiał, że innowacyjność nie polega wyłącznie na generowaniu pomysłów, lecz przede wszystkim na ich skutecznym wdrażaniu oraz osiąganiu konkretnych rezultatów organizacyjnych i ekonomicznych.

1.4. Klasyfikacja innowacji

Jednym z najczęściej przywoływanych podziałów innowacji jest klasyfikacja według przedmiotu zmiany. W tym ujęciu wyróżnia się cztery podstawowe rodzaje innowacji: **produktowe, procesowe, organizacyjne i marketingowe.** Taki podział pojawia się zarówno w literaturze naukowej, jak i w opracowaniach odwołujących się do Podręcznika Oslo OECD.

- **Innowacje produktowe** - polegają na wprowadzeniu nowego produktu lub istotnym ulepszeniu już istniejącego wyrobu albo usługi. Zmiana może dotyczyć parametrów technicznych, funkcjonalności, trwałości, bezpieczeństwa, ergonomii czy materiałów wykorzystanych do produkcji. W realiach przedsiębiorstw produkcyjnych może to oznaczać na przykład opracowanie bardziej energooszczędnego urządzenia, zastosowanie nowego materiału konstrukcyjnego albo zwiększenie niezawodności produktu końcowego.
- **Innowacje procesowe** - odnoszą się do sposobu wytwarzania, dostarczania i realizacji działań w przedsiębiorstwie. Ich celem jest zwykle zwiększenie efektywności procesu, poprawa jakości, zmniejszenie zużycia zasobów lub skrócenie czasu realizacji. W literaturze wskazuje się, że obejmują one wdrożenie nowych albo znacząco ulepszonych metod produkcji lub dostaw, a także zmiany dotyczące projektowania,

zaopatrzenia czy przepływu informacji. Dla studentów automatyki i robotyki ten rodzaj innowacji jest szczególnie istotny, ponieważ często wiąże się z modernizacją stanowisk pracy, automatyzacją operacji, zastosowaniem systemów sterowania, robotyzacją albo cyfryzacją nadzoru nad procesem.

- **Innowacje organizacyjne -** obejmują zmiany zachodzące w strukturach, procedurach, metodach zarządzania oraz organizacji miejsca pracy. Mogą dotyczyć sposobu planowania produkcji, podziału kompetencji, komunikacji między działami, organizacji utrzymania ruchu czy zarządzania jakością. Choć nie zawsze są tak widoczne jak zmiany technologiczne, bardzo często stanowią warunek skutecznego wdrożenia rozwiązań technicznych.
- **Innowacje marketingowe** - dotyczą zmian w sposobie projektowania i prezentacji produktu, jego promocji, polityki cenowej czy kanałów dystrybucji. W przedsiębiorstwach przemysłowych mogą wydawać się mniej istotne niż innowacje procesowe, jednak w praktyce wpływają na pozycję rynkową firmy, jej zdolność do dotarcia do klienta oraz budowania przewagi konkurencyjnej (Trzepizur 2016).

Poza tym podziałem w literaturze wyróżnia się także innowacje **przełomowe i przyrostowe**.

- **Innowacje przełomowe** mają charakter fundamentalny i prowadzą do wyraźnej zmiany sposobu działania przedsiębiorstwa lub rynku.
- **Innowacje przyrostowe** polegają na stopniowym udoskonalaniu już istniejących produktów, procesów lub metod organizacji. W praktyce przedsiębiorstw produkcyjnych właśnie innowacje przyrostowe występują bardzo często, ponieważ pozwalają osiągać realne korzyści bez konieczności ponoszenia tak wysokiego ryzyka jak w przypadku zmian radykalnych (Kubiński 2017).

1.5. Usprawnienia jako element ciągłego doskonalenia przedsiębiorstwa

W analizie funkcjonowania przedsiębiorstw nie można ograniczać się wyłącznie do pojęcia innowacji. Równie ważną kategorią są usprawnienia, które najczęściej dotyczą doskonalenia istniejących procesów, organizacji pracy i przepływu wartości. Usprawnienie oznacza celową zmianę prowadzącą do poprawy określonego parametru działania przedsiębiorstwa. Może to być skrócenie czasu realizacji operacji, obniżenie kosztu

jednostkowego, zmniejszenie liczby reklamacji, poprawa bezpieczeństwa, wzrost wykorzystania maszyn czy uproszczenie obiegu informacji.

Z perspektywy dydaktycznej usprawnienia są szczególnie ważne, ponieważ uczą studentów patrzenia na przedsiębiorstwo w sposób analityczny. Zamiast pytać jedynie, czy firma jest „nowoczesna", student powinien umieć wskazać, które elementy jej funkcjonowania wymagają poprawy, jakie są źródła niesprawności i jakie działania można wdrożyć, aby osiągnąć lepsze wyniki. Taki kierunek myślenia został wyraźnie zapisany w sylabusie, gdzie jednym z celów kształcenia jest umiejętność stosowania metod i narzędzi pozwalających na identyfikację problemów występujących w przedsiębiorstwie oraz metod ich eliminacji.
Warto podkreślić, że usprawnienie nie musi oznaczać kosztownej inwestycji. Czasami duży efekt można osiągnąć dzięki lepszemu rozplanowaniu stanowiska pracy, zmianie kolejności działań, redukcji zbędnych czynności kontrolnych, poprawie komunikacji między działami albo standaryzacji wybranych operacji. W tym sensie usprawnienia są bliskie idei ciągłego doskonalenia, która zakłada, że organizacja powinna systematycznie poszukiwać i wdrażać nawet drobne zmiany prowadzące do poprawy funkcjonowania.

1.6. Znaczenie pracowników w procesie innowacji i usprawnień

Współczesne przedsiębiorstwo nie jest innowacyjne wyłącznie dzięki wyposażeniu technicznemu. Równie istotny jest czynnik ludzki. Sylabus przedmiotu wskazuje bezpośrednio na rolę pracowników przedsiębiorstwa w ciągłym doskonaleniu procesów. Jest to założenie bardzo trafne, ponieważ to właśnie pracownicy najczęściej obserwują codzienne problemy występujące w procesach, dostrzegają marnotrawstwo, przestoje, powtarzające się błędy i zbędne operacje.

W literaturze zwraca się uwagę, że innowacje rozpoczynają się od pomysłów i idei, które następnie są rozwijane oraz wdrażane w organizacji. Kreatywność stanowi więc istotny warunek innowacyjności, ale sama kreatywność nie wystarcza. Potrzebne są jeszcze warunki organizacyjne umożliwiające ocenę pomysłu, jego dopracowanie i praktyczne wdrożenie. Oznacza to, że przedsiębiorstwo powinno nie tylko oczekiwać inicjatywy od pracowników, ale także tworzyć mechanizmy sprzyjające zgłaszaniu i testowaniu usprawnień (Wiśniewska, Janasz 2023).

W przypadku firm produkcyjnych szczególnie cenne są kompetencje pracowników związane z obserwacją procesu, analizą przyczyn problemów oraz oceną skutków

proponowanych zmian. Dlatego w kształceniu studentów ważne jest rozwijanie nie tylko wiedzy teoretycznej o innowacjach, lecz także umiejętności praktycznych, takich jak analiza procesu, opis strumienia wartości, przygotowanie mapy procesu czy zaproponowanie usprawnienia na podstawie zidentyfikowanych niesprawności. Te elementy również zostały przewidziane w części projektowej sylabusa.

1.7. Innowacje i usprawnienia w praktyce przedsiębiorstw produkcyjnych

Na kierunkach technicznych, w tym na automatyce i robotyce, szczególnie istotne jest praktyczne rozumienie innowacji i usprawnień. Student powinien potrafić odnieść te pojęcia do realnych zjawisk obserwowanych w przedsiębiorstwie. W warunkach produkcyjnych innowacja produktowa może oznaczać opracowanie nowego podzespołu lub zmianę konstrukcji poprawiającą funkcjonalność wyrobu. Innowacja procesowa może polegać na robotyzacji stanowiska, wdrożeniu systemu wizyjnego, zmianie organizacji przepływu materiału albo zastosowaniu cyfrowego monitorowania parametrów procesu. Innowacja organizacyjna może przyjąć postać nowego sposobu planowania zleceń, integracji działu produkcji z utrzymaniem ruchu czy wprowadzenia standardów raportowania odchyleń.

Z kolei usprawnienia często mają bardziej lokalny, ale bardzo praktyczny charakter. Mogą obejmować zmianę układu stanowiska montażowego, ograniczenie zbędnych przemieszczeń operatora, poprawę oznaczeń materiałów, skrócenie czasu przezbrojenia lub uporządkowanie kolejności operacji. Tego rodzaju działania nie zawsze są medialne, ale w rzeczywistych warunkach przedsiębiorstwa właśnie one często decydują o wzroście wydajności, jakości i stabilności procesu.

Z tego względu przedmiot „Innowacje i usprawnienia w firmach” słusznie łączy problematykę innowacji z analizą strumienia wartości, analizą procesów oraz projektowaniem usprawnień. Student ma nie tylko rozumieć znaczenie innowacji, ale także umieć zastosować metody analizy techniczno-ekonomicznej oraz wskazać sposoby usprawniania procesów i strumienia wartości.

2. Analiza procesów i strumienia wartości jako podstawa identyfikacji usprawnień w przedsiębiorstwie

Współczesne przedsiębiorstwa coraz częściej postrzegane są jako złożone systemy procesów, których efektywność decyduje o ich konkurencyjności oraz zdolności do wdrażania

innowacji. W tym kontekście kluczowe znaczenie ma podejście procesowe, które umożliwia identyfikację zależności między działaniami realizowanymi w organizacji oraz ocenę ich wpływu na tworzenie wartości dla klienta. Zgodnie z założeniami przedmiotu „Innowacje i usprawnienia w firmach", szczególny nacisk kładziony jest na analizę procesów oraz strumienia wartości, a także identyfikację problemów i proponowanie usprawnień. Umiejętność ta stanowi fundament działań inżynierskich i menedżerskich, ponieważ pozwala przejść od ogólnego rozumienia funkcjonowania przedsiębiorstwa do konkretnych działań doskonalących.

2.1. Podejście procesowe w zarządzaniu przedsiębiorstwem

Podejście procesowe polega na analizie organizacji jako zbioru powiązanych ze sobą procesów, które przekształcają zasoby wejściowe (materiały, informacje, energię) w produkty lub usługi stanowiące wartość dla klienta. W odróżnieniu od podejścia funkcjonalnego, koncentrującego się na strukturze organizacyjnej, podejście procesowe skupia się na przebiegu działań oraz ich efektywności. Proces można zdefiniować jako uporządkowany ciąg działań prowadzących do osiągnięcia określonego rezultatu. Każdy proces posiada wejścia, wyjścia, zasoby oraz właściciela odpowiedzialnego za jego realizację (Hammer, Champy 1993). W praktyce przedsiębiorstwa wyróżnia się procesy podstawowe (bezpośrednio tworzące wartość dla klienta) oraz procesy pomocnicze (wspierające działalność podstawową). Znaczenie podejścia procesowego wynika z faktu, że wiele problemów w przedsiębiorstwie nie wynika z błędów pojedynczych pracowników, lecz z niewłaściwej organizacji procesu. Dlatego analiza procesów stanowi punkt wyjścia do identyfikacji nieefektywności oraz projektowania usprawnień.

2.2. Istota i znaczenie strumienia wartości

Jednym z kluczowych pojęć wykorzystywanych w analizie procesów jest strumień wartości (*value stream*). Oznacza on wszystkie działania, zarówno dodające wartość, jak i niedodające wartości, które są niezbędne do przekształcenia surowca w gotowy produkt lub usługę (Rother, Shook 2003).

Analiza strumienia wartości pozwala odpowiedzieć na fundamentalne pytania:

- Które działania rzeczywiście tworzą wartość dla klienta?

- Które działania są zbędne lub mogą zostać ograniczone?
- Gdzie występują opóźnienia, przestoje i marnotrawstwo?

W literaturze podkreśla się, że zdolność do identyfikacji i eliminacji działań niedodających wartości stanowi jeden z kluczowych elementów doskonalenia procesów oraz zwiększania efektywności przedsiębiorstwa (Womack, Jones 2001). W kontekście sylabusa przedmiotu szczególnie istotna jest umiejętność opisu strumienia wartości oraz jego analizy, co bezpośrednio przekłada się na zdolność do projektowania usprawnień.

2.3. Mapowanie strumienia wartości (VSM)

Jednym z najważniejszych narzędzi analizy procesów jest mapowanie strumienia wartości (*Value Stream Mapping* - VSM). Metoda ta polega na graficznym przedstawieniu przepływu materiałów i informacji w procesie produkcyjnym lub usługowym.

Mapa strumienia wartości obejmuje m.in.:

- kolejne etapy procesu,
- czasy realizacji operacji,
- zapasy międzyoperacyjne,
- przepływ informacji,
- miejsca występowania opóźnień i wąskich gardeł.

Celem VSM jest uzyskanie pełnego obrazu procesu oraz identyfikacja obszarów wymagających usprawnień. W praktyce wyróżnia się dwa podstawowe typy map:

- stan obecny (*as-is*) - przedstawiający rzeczywisty przebieg procesu,
- stan przyszły (*to-be*) - uwzględniający proponowane usprawnienia.

Zastosowanie VSM umożliwia nie tylko analizę procesu, ale również projektowanie jego optymalnej struktury. Narzędzie to jest szeroko wykorzystywane w przedsiębiorstwach produkcyjnych oraz stanowi istotny element koncepcji Lean Management.

2.4. Identyfikacja problemów w procesach

Kluczowym etapem analizy procesów jest identyfikacja problemów oraz tzw. niesprawności (marnotrawstwa). W literaturze Lean wyróżnia się kilka podstawowych rodzajów marnotrawstwa, takich jak nadprodukcja, zbędny transport, oczekiwanie, nadmierne zapasy czy błędy jakościowe (Ohno 1988).

Problemy w procesach mogą mieć różnorodne źródła, m.in.:

- niewłaściwą organizację stanowiska pracy,
- brak standaryzacji działań,
- niedostosowanie technologii do wymagań procesu,
- błędy w przepływie informacji,
- niewystarczające kompetencje pracowników.

Zgodnie z sylabusem przedmiotu, student powinien potrafić nie tylko zidentyfikować problem, ale również zaproponować sposób jego rozwiązania. W praktyce oznacza to konieczność analizy przyczyn problemów, a nie jedynie ich objawów.

Do najczęściej stosowanych metod identyfikacji przyczyn należą:

- diagram Ishikawy (przyczynowo-skutkowy),
- metoda 5 *Why* (5 razy „dlaczego”),
- analiza Pareto.

2.5. Analiza procesów jako podstawa usprawnień

Analiza procesów stanowi fundament projektowania usprawnień w przedsiębiorstwie. Jej celem jest określenie, które elementy procesu wymagają zmiany oraz jakie działania należy podjąć w celu poprawy jego efektywności.
W praktyce analiza ta obejmuje:

- ocenę czasu realizacji poszczególnych operacji,
- identyfikację wąskich gardeł,

- analizę przepływu materiałów i informacji,
- ocenę wykorzystania zasobów.

Na tej podstawie możliwe jest zaproponowanie usprawnień, takich jak:

- skrócenie czasu realizacji procesu,
- eliminacja zbędnych operacji,
- poprawa organizacji pracy,
- automatyzacja wybranych etapów,
- lepsze wykorzystanie zasobów.

W literaturze podkreśla się, że skuteczne usprawnienia powinny być oparte na danych oraz analizie rzeczywistych procesów, a nie jedynie na intuicji (Davenport 1993).

2.6. Projektowanie usprawnień w przedsiębiorstwach

Projektowanie usprawnień jest procesem wymagającym zarówno wiedzy teoretycznej, jak i umiejętności praktycznych. Obejmuje ono:

- analizę stanu obecnego,
- identyfikację problemów,
- opracowanie propozycji zmian,
- ocenę ich wykonalności,
- wdrożenie i kontrolę efektów.

W kontekście zajęć projektowych, opisanych w sylabusie, student powinien przejść przez wszystkie te etapy, wykorzystując dane z rzeczywistego przedsiębiorstwa. Szczególnie istotne jest przygotowanie projektu usprawnień oraz jego uzasadnienie na podstawie przeprowadzonej analizy. W praktyce przedsiębiorstw usprawnienia mogą dotyczyć zarówno pojedynczych operacji, jak i całych procesów. Ich skuteczność zależy od właściwego dopasowania do specyfiki organizacji oraz zaangażowania pracowników.

Analiza procesów oraz strumienia wartości stanowi podstawę identyfikacji usprawnień w przedsiębiorstwie. Podejście procesowe pozwala na kompleksowe spojrzenie na działalność organizacji oraz zrozumienie zależności między poszczególnymi działaniami. Narzędzia takie jak mapowanie strumienia wartości umożliwiają identyfikację problemów, marnotrawstwa oraz obszarów wymagających doskonalenia. Na tej podstawie możliwe jest projektowanie usprawnień prowadzących do zwiększenia efektywności, poprawy jakości oraz redukcji kosztów. Z perspektywy kształcenia inżynierskiego szczególnie istotne jest rozwijanie umiejętności analizy procesów oraz projektowania usprawnień w oparciu o rzeczywiste dane. Takie podejście pozwala studentom przygotować się do pracy w przedsiębiorstwach, w których ciągłe doskonalenie stanowi podstawowy element funkcjonowania.

3. Metody i narzędzia usprawnień procesów w przedsiębiorstwie

Wdrażanie innowacji oraz usprawnień w przedsiębiorstwie nie jest możliwe bez zastosowania odpowiednich metod i narzędzi analitycznych oraz organizacyjnych. O ile w poprzednim rozdziale wskazano znaczenie analizy procesów i strumienia wartości, o tyle w niniejszym rozdziale skoncentrowano się na praktycznych rozwiązaniach wspierających doskonalenie działalności przedsiębiorstw. Zgodnie z założeniami sylabusa przedmiotu „Innowacje i usprawnienia w firmach", studenci powinni nabyć umiejętność stosowania narzędzi pozwalających na identyfikację problemów oraz ich eliminację, a także projektowania usprawnień procesów. W praktyce oznacza to konieczność poznania metod, które znajdują zastosowanie zarówno w przedsiębiorstwach produkcyjnych, jak i usługowych. Współczesne podejścia do doskonalenia procesów opierają się w dużej mierze na koncepcjach Lean Management, zarządzania jakością oraz podejściu procesowym. Narzędzia te pozwalają nie tylko na identyfikację problemów, ale również na systematyczne ich rozwiązywanie oraz zapobieganie ich ponownemu występowaniu.

3.1. Koncepcja Lean Management jako podstawa usprawnień

Lean Management stanowi jedną z najważniejszych koncepcji wykorzystywanych w doskonaleniu procesów przedsiębiorstwa. Jej głównym celem jest eliminacja marnotrawstwa oraz maksymalizacja wartości dostarczanej klientowi (Womack, Jones 2001).

W podejściu Lean kluczowe jest rozróżnienie działań:

- dodających wartość (*Value Added*),
- niedodających wartości, ale koniecznych,
- niedodających wartości (marnotrawstwo).

Podstawowe zasady Lean obejmują:

- identyfikację wartości dla klienta,
- analizę strumienia wartości,
- zapewnienie płynności przepływu,
- stosowanie systemu ssącego (*pull*),
- ciągłe doskonalenie (*kaizen*) (Womack, Jones 2001).

Zastosowanie Lean Management pozwala na systematyczne doskonalenie procesów oraz eliminację działań nieefektywnych, co bezpośrednio przekłada się na poprawę wyników przedsiębiorstwa.

3.2. Marnotrawstwo w procesach - identyfikacja i eliminacja

Jednym z kluczowych elementów Lean Management jest identyfikacja marnotrawstwa (muda). Koncepcja ta została rozwinięta w ramach Systemu Produkcyjnego Toyoty i obejmuje siedem podstawowych rodzajów strat (Ohno 1988):

1) nadprodukcja,
2) oczekiwanie,
3) zbędny transport,
4) nadmierne przetwarzanie,
5) zapasy,
6) zbędny ruch,
7) błędy i poprawki.

W późniejszych opracowaniach dodano również ósmy rodzaj marnotrawstwa - niewykorzystany potencjał pracowników (Liker 2004). Eliminacja marnotrawstwa stanowi podstawę poprawy efektywności procesów i może prowadzić do:

- skrócenia czasu realizacji,
- redukcji kosztów,
- poprawy jakości,
- zwiększenia wydajności (Womack, Jones 2001).

3.3. Metoda 5S jako narzędzie organizacji pracy

Metoda 5S jest jednym z podstawowych narzędzi Lean Management, wykorzystywanym do organizacji stanowiska pracy. Jej nazwa pochodzi od pięciu japońskich terminów:

- Seiri - selekcja,
- Seiton - systematyka,
- Seiso - sprzątanie,
- Seiketsu - standaryzacja,
- Shitsuke - samodyscyplina (Hirano 1995).

Celem metody 5S jest stworzenie uporządkowanego, bezpiecznego i efektywnego środowiska pracy. Wdrożenie tej metody prowadzi do poprawy ergonomii, zmniejszenia liczby błędów oraz zwiększenia wydajności pracy. W praktyce przedsiębiorstw produkcyjnych metoda 5S stanowi często pierwszy etap wdrażania koncepcji Lean.

3.4. Kaizen - ciągłe doskonalenie

Kaizen oznacza filozofię ciągłego doskonalenia, polegającą na systematycznym wprowadzaniu drobnych usprawnień w procesach przedsiębiorstwa (Imai 1986).

Podstawowe założenia kaizen obejmują:

- zaangażowanie wszystkich pracowników,

- koncentrację na małych zmianach,
- ciągłe poszukiwanie usprawnień.

W literaturze podkreśla się, że kaizen sprzyja budowaniu kultury organizacyjnej opartej na innowacyjności oraz uczeniu się (Imai 1986). W odróżnieniu od innowacji przełomowych, podejście to zakłada stopniowe doskonalenie procesów.

3.5. Analiza wartości w procesie usprawnień

Analiza wartości stanowi narzędzie pozwalające na ocenę relacji między funkcją produktu lub procesu a jego kosztem. Jej celem jest osiągnięcie wymaganej funkcjonalności przy minimalnych nakładach (Miles 1961).

Proces analizy wartości obejmuje:

- identyfikację funkcji,
- określenie kosztów realizacji funkcji,
- ocenę efektywności,
- poszukiwanie alternatywnych rozwiązań.

Zastosowanie analizy wartości umożliwia:

- redukcję kosztów,
- uproszczenie konstrukcji,
- poprawę efektywności procesów.

Zgodnie z sylabusem, analiza wartości stanowi jeden z kluczowych elementów kształcenia.

3.6. Narzędzia analizy przyczyn problemów

Skuteczne usprawnienia wymagają identyfikacji przyczyn problemów. Do najczęściej stosowanych narzędzi należą:

- **diagram Ishikawy** - analiza przyczynowo-skutkowa (Ishikawa 1985),

- **metoda 5 Why** - identyfikacja przyczyn źródłowych (Ohno 1988),
- **analiza Pareto** - identyfikacja najważniejszych problemów (Juran 1988).

Zastosowanie tych narzędzi pozwala na eliminację rzeczywistych przyczyn problemów, a nie tylko ich objawów.

3.7. Automatyzacja i cyfryzacja procesów

Współczesne przedsiębiorstwa coraz częściej wykorzystują technologie cyfrowe jako narzędzie usprawnień. Automatyzacja procesów produkcyjnych oraz wykorzystanie systemów informatycznych pozwalają na zwiększenie efektywności i jakości działania (Kubiński 2017).

Do najważniejszych rozwiązań należą:

- robotyzacja procesów,
- systemy ERP i MES,
- analiza danych produkcyjnych,
- Internet Rzeczy (IoT).

Zastosowanie tych technologii umożliwia lepsze zarządzanie procesami oraz szybsze reagowanie na problemy.

3.8. Wdrażanie usprawnień w przedsiębiorstwie

Proces wdrażania usprawnień obejmuje kilka etapów:

- analizę stanu obecnego,
- identyfikację problemów,
- opracowanie rozwiązań,
- wdrożenie,
- kontrolę efektów (Davenport 1993).

W praktyce kluczowe znaczenie ma zaangażowanie pracowników oraz odpowiednie zarządzanie zmianą. Nie wszystkie usprawnienia kończą się sukcesem, dlatego istotne jest monitorowanie efektów oraz ciągłe doskonalenie.

Metody i narzędzia usprawnień stanowią podstawę skutecznego zarządzania procesami w przedsiębiorstwie. Ich zastosowanie umożliwia identyfikację problemów, eliminację marnotrawstwa oraz poprawę efektywności działania organizacji. Z perspektywy kształcenia inżynierskiego kluczowe jest rozwijanie umiejętności praktycznego wykorzystania tych narzędzi. Studenci powinni nie tylko znać ich teoretyczne podstawy, ale również potrafić zastosować je w analizie rzeczywistych procesów.

4. Studium przypadku: usprawnienie zautomatyzowanej linii sortowania i pakowania komponentów przemysłowych

4.1. Charakterystyka przedsiębiorstwa i procesu

Analizowane przedsiębiorstwo to średniej wielkości firma produkcyjna działająca w branży elektrotechnicznej, zajmująca się wytwarzaniem komponentów do systemów automatyki przemysłowej. W strukturze organizacyjnej firmy funkcjonuje dział produkcji seryjnej, w którym wykorzystywane są zautomatyzowane linie technologiczne.

Przedmiotem analizy jest fragment procesu obejmujący:

- sortowanie komponentów po obróbce mechanicznej,
- kontrolę jakości (wizualną i pomiarową),
- pakowanie wyrobów gotowych.

Proces realizowany jest na zautomatyzowanej linii wyposażonej w:

- przenośniki taśmowe,
- system wizyjny,
- stanowisko robota przemysłowego (*pick-and-place*),
- stanowisko pakowania półautomatycznego.

Zgodnie z podejściem procesowym, analizowany układ stanowi część strumienia wartości przedsiębiorstwa, którego celem jest dostarczenie klientowi produktu o wymaganej jakości.

4.2. Opis procesu - stan obecny (AS-IS)

Proces przebiega w następujących etapach:

- **Dostarczenie komponentów**
  Elementy po obróbce trafiają na przenośnik wejściowy.
- **Sortowanie wstępne**
  Operator nadzoruje przepływ detali i usuwa widoczne uszkodzenia.
- **Kontrola wizyjna**
  System kamer identyfikuje wady powierzchniowe.
- **Odbiór przez robota**
  Robot przemysłowy pobiera elementy i przekłada je na stanowisko pakowania.
- **Pakowanie**
  Operator umieszcza elementy w opakowaniach zbiorczych.
- **Transport do magazynu**
  Gotowe pakiety trafiają na linię transportową.

Czas całkowity realizacji procesu dla jednej partii wynosi średnio 18 min, przy czym analiza wykazała duże zróżnicowanie czasów poszczególnych operacji.

4.3. Identyfikacja problemów w procesie

Na podstawie obserwacji procesu, analizy czasów oraz rozmów z pracownikami zidentyfikowano następujące problemy:

1) **Wąskie gardło - stanowisko pakowania**

Największe opóźnienia występują na etapie pakowania. Operator nie nadąża z odbiorem elementów dostarczanych przez robota, co prowadzi do:

- zatrzymań robota,

- kumulacji detali na przenośniku.

Zjawisko to odpowiada klasycznemu problemowi wąskiego gardła w procesie (Goldratt 1990).

2) **Niska efektywność systemu wizyjnego**

System wizyjny generuje dużą liczbę błędnych odrzuceń (*false reject*), co skutkuje:

- koniecznością ręcznej weryfikacji,
- wydłużeniem czasu procesu.

Problemy te wynikają z niedostatecznej kalibracji systemu oraz zmiennych warunków oświetleniowych.

3) **Przestoje robota przemysłowego**

Robot często zatrzymuje się z powodu:

- braku detali (przerwy w dostawie),
- przepełnienia stanowiska pakowania.

Powoduje to spadek wykorzystania jego potencjału.

4) **Nadmierne zapasy międzyoperacyjne**

Na odcinku między kontrolą a pakowaniem obserwuje się nagromadzenie elementów, co stanowi przykład marnotrawstwa zapasów (Ohno 1988).

5) **Niewykorzystany potencjał pracowników**

Operatorzy wykonują powtarzalne czynności manualne, które mogłyby zostać zautomatyzowane.

4.4. Analiza przyczyn problemów

W celu identyfikacji przyczyn źródłowych zastosowano metodę 5 Why (Ohno 1988).

Przykład - problem opóźnień w pakowaniu:

Dlaczego występują opóźnienia?
→ Operator nie nadąża z pakowaniem

Dlaczego nie nadąża?

→ Tempo pracy robota jest wyższe niż tempo operatora

Dlaczego tempo nie jest dostosowane?

→ Brak synchronizacji systemu

Dlaczego brak synchronizacji?

→ Brak sprzężenia zwrotnego między stanowiskami

Dlaczego go nie ma?

→ System nie został zaprojektowany jako zintegrowany

➞ Przyczyna źródłowa: brak integracji systemu sterowania

4.5. Propozycje usprawnień (TO-BE)

Na podstawie analizy zaproponowano następujące usprawnienia:

1) **Automatyzacja stanowiska pakowania**

Zastosowanie drugiego robota (lub cobota) do pakowania pozwoli na:

- eliminację wąskiego gardła,
- zwiększenie wydajności procesu.

Automatyzacja procesów jest jednym z kluczowych kierunków usprawnień w nowoczesnych przedsiębiorstwach (Kubiński 2017).

2) **Integracja systemów sterowania**

Wprowadzenie systemu sterowania nadrzędnego (np. PLC + MES) umożliwi:

- synchronizację pracy robota i pakowania,
- eliminację przestojów.

3) **Optymalizacja systemu wizyjnego**

Działania:

- kalibracja kamer,

- poprawa oświetlenia,
- wykorzystanie algorytmów AI.

Pozwoli to ograniczyć liczbę błędnych odrzuceń.

4) **Wprowadzenie buforów sterowanych**

Zastosowanie kontrolowanych buforów międzyoperacyjnych:

- zmniejszy ryzyko zatrzymań,
- poprawi płynność przepływu.

5) **Wdrożenie zasad Lean**

- redukcja zapasów,
- eliminacja zbędnych operacji,
- poprawa organizacji stanowisk pracy (5S).

Lean Management umożliwia systematyczne doskonalenie procesów (Womack, Jones 2001).

4.6. Ocena efektów usprawnień

Po wdrożeniu usprawnień przewiduje się:

- skrócenie czasu procesu o ok. **30–40%**,
- redukcję przestojów robota o **60%**,
- zmniejszenie liczby błędów jakościowych,
- zwiększenie wydajności linii.

Zgodnie z podejściem inżynierskim, każda zmiana powinna być monitorowana i oceniana na podstawie danych (Davenport 1993).

4.7. Wnioski

Studium przypadku pokazuje, że:

- problemy w procesach często wynikają z braku integracji systemów,
- automatyzacja nie zawsze oznacza pełną efektywność,
- kluczowe znaczenie ma analiza procesu i identyfikacja wąskich gardeł.

Z perspektywy studentów kierunku automatyka i robotyka szczególnie ważne jest:

- rozumienie zależności między systemami technicznymi,
- umiejętność analizy procesów,
- projektowanie usprawnień z wykorzystaniem technologii.

Zadanie 1. Identyfikacja marnotrawstwa

Polecenie: Wskaż minimum 5 rodzajów marnotrawstwa występujących w analizowanej linii.

Skorzystaj z klasyfikacji:

- nadprodukcja
- oczekiwanie
- transport
- zapasy
- błędy
- ruch
- nadmierne przetwarzanie

Wskazówka: Odnieś się do sytuacji takich jak przestoje robota czy kolejki przed pakowaniem.

Przykładowe rozwiązanie

W analizowanym procesie można wskazać następujące rodzaje marnotrawstwa:

- **Oczekiwanie**: występuje wtedy, gdy robot zatrzymuje się z powodu przepełnienia stanowiska pakowania albo operator czeka na ręczną weryfikację błędnie odrzuconych detali. Jest to klasyczny przykład strat czasu.
- **Zapasy**: między kontrolą wizyjną a pakowaniem pojawiają się nagromadzone komponenty oczekujące na dalszą obróbkę. Oznacza to tworzenie zapasów międzyoperacyjnych, które nie zwiększają wartości wyrobu.
- **Nadmierne przetwarzanie**: ręczna weryfikacja elementów odrzuconych przez system wizyjny oznacza dodatkową czynność, która w dobrze zaprojektowanym procesie powinna być ograniczona.
- **Błędy i poprawki**: błędne odrzucenia przez system wizyjny powodują konieczność dodatkowej kontroli. To przykład strat jakościowych.
- **Niewykorzystany potencjał pracowników**: operator wykonuje powtarzalne czynności pakowania, które częściowo mogłyby zostać zautomatyzowane. Oznacza to, że jego potencjał nie jest wykorzystywany do działań bardziej wartościowych.
- **Zbędny ruch**: jeżeli operator musi stale przemieszczać się między odkładaniem elementów, kontrolą opakowań i organizacją stanowiska, pojawia się dodatkowy ruch niedodający wartości.
- **Nadprodukcja w ujęciu lokalnym**: robot działa szybciej niż stanowisko pakowania, co powoduje nadmierny napływ detali do kolejnego etapu procesu. Nie jest to nadprodukcja w sensie rynkowym, ale lokalne przeciążenie procesu.

Wniosek: Najbardziej widoczne marnotrawstwa w tym procesie to oczekiwanie, zapasy, błędy, nadmierne przetwarzanie oraz niewykorzystany potencjał pracowników. Tego rodzaju analiza jest zgodna z podejściem Lean, które koncentruje się na eliminacji czynności niedodających wartości.

Zadanie 2. Analiza wąskiego gardła

Polecenie: Które stanowisko stanowi wąskie gardło procesu? Jakie są jego konsekwencje?

Odpowiedz na pytania:

1) Które stanowisko stanowi wąskie gardło procesu?
2) Jakie są jego konsekwencje dla całej linii?

3) Jak wpływa ono na:

- czas realizacji procesu,
- wykorzystanie robota,
- poziom zapasów?

Przykładowe rozwiązanie

Wąskim gardłem procesu jest stanowisko pakowania manualnego.

Stanowisko to ma najdłuższy czas operacji, wynoszący 180 sekund, podczas gdy pozostałe etapy procesu są realizowane szybciej. Zgodnie z zasadą analizy ograniczeń, wydajność całego procesu jest wyznaczana przez etap najwolniejszy, czyli właśnie przez pakowanie (Goldratt 1990).

Konsekwencje występowania wąskiego gardła:

- **Wydłużenie czasu realizacji całego procesu**: cała linia pracuje w tempie najwolniejszego stanowiska.
- **Powstawanie zapasów międzyoperacyjnych**: detale odkładają się przed stanowiskiem pakowania.
- **Przestoje robota**: robot okresowo zatrzymuje się, ponieważ nie ma możliwości przekazania kolejnych elementów.
- **Niskie wykorzystanie potencjału automatyzacji**: szybkie stanowiska nie pracują zgodnie ze swoimi możliwościami.
- **Zaburzenie płynności przepływu**: proces nie ma charakteru ciągłego, lecz nierównomierny.

Wniosek: Wąskie gardło powoduje, że nawet nowoczesne elementy linii, takie jak robot przemysłowy, nie przynoszą pełnych korzyści. To pokazuje, że automatyzacja pojedynczego etapu bez synchronizacji całego procesu nie gwarantuje wysokiej efektywności.

Zadanie 3. Analiza przyczyn problemu metodą 5 Why

Polecenie: Wybierz problem i wykonaj analizę 5 Why. Wybierz jeden problem: przestoje robota lub opóźnienia pakowania

Wykonaj analizę metodą **5 Why** i zapisz:

- problem główny
- kolejne przyczyny
- przyczynę źródłową

Przykładowe rozwiązanie dla problemu: opóźnienia pakowania

Problem: Występują opóźnienia na stanowisku pakowania.

Dlaczego 1? Bo operator nie nadąża z pakowaniem elementów.

Dlaczego 2? Bo elementy trafiają do pakowania szybciej, niż mogą być ręcznie umieszczane w opakowaniach.

Dlaczego 3? Bo robot pick-and-place pracuje znacznie szybciej niż operator.

Dlaczego 4? Bo nie ma synchronizacji między tempem pracy robota a tempem pakowania.

Dlaczego 5? Bo system sterowania nie został zaprojektowany jako zintegrowany układ z informacją zwrotną.

Przyczyna źródłowa: brak integracji i synchronizacji stanowisk w systemie sterowania.

Wniosek: przyczyną problemu nie jest wyłącznie „zbyt wolny operator", ale przede wszystkim nieprawidłowo zaprojektowany układ procesu. To bardzo ważne dydaktycznie, bo pokazuje studentom różnicę między objawem a przyczyną źródłową.

Zadanie 4. Propozycje usprawnień

Polecenie: Zaproponuj minimum **trzy usprawnienia procesu**, które:

- zmniejszą czas realizacji,
- poprawią płynność przepływu,
- ograniczą błędy.

Każde usprawnienie opisz:

- co zmieniasz?
- jaki problem rozwiązujesz?
- jaki efekt uzyskasz?

Przykładowe rozwiązanie

- **Automatyzacja pakowania:** na stanowisku pakowania można zastosować cobota lub drugiego robota. Rozwiązywany problem: wąskie gardło procesu. **Efekt:** skrócenie czasu pakowania, zmniejszenie przestojów i poprawa wydajności.
- **Integracja systemu sterowania**: należy wprowadzić sprzężenie zwrotne między robotem a stanowiskiem pakowania, np. przez sterownik PLC lub system MES. Rozwiązywany problem: brak synchronizacji. **Efekt:** płynniejszy przepływ materiału i ograniczenie zatrzymań.
- **Poprawa działania systemu wizyjnego**: należy przeprowadzić kalibrację kamer, poprawić oświetlenie i dopracować parametry kontroli. Rozwiązywany problem: błędne odrzucenia detali. **Efekt:** mniej błędów, mniej pracy ręcznej, krótszy czas procesu.
- **Wdrożenie 5S na stanowisku pakowania**: uporządkowanie stanowiska, standaryzacja odkładania materiałów i poprawa ergonomii. Rozwiązywany problem: zbędny ruch operatora. **Efekt:** skrócenie czasu czynności ręcznych.

Wniosek: najlepsze usprawnienia to te, które jednocześnie poprawiają przepływ, redukują błędy i zwiększają wykorzystanie zasobów.

Zadanie 5. Czy pełna automatyzacja linii zawsze jest najlepszym rozwiązaniem?

Uzasadnij:

- technicznie
- ekonomicznie

Przykładowe rozwiązanie

Pełna automatyzacja nie zawsze jest najlepszym rozwiązaniem. Z technicznego punktu widzenia może zwiększyć wydajność i powtarzalność, ale z ekonomicznego punktu widzenia wymaga wysokich nakładów inwestycyjnych. W niektórych procesach bardziej opłacalne może być częściowe zautomatyzowanie wybranych etapów oraz lepsza organizacja pracy. W analizowanym przypadku widać, że sam robot pick-and-place nie rozwiązał problemów procesu, ponieważ pakowanie pozostało ręczne. Oznacza to, że automatyzacja jednego stanowiska bez integracji z pozostałymi etapami może prowadzić do nowych niesprawności.

Dlatego decyzja o automatyzacji powinna być poprzedzona analizą procesu, kosztów i spodziewanych efektów.

Zadanie 6. Dane wyjściowe do analizy

Tabela 1. Dane procesu

Etap procesu	Czas operacji (s)	Wydajność (szt./min)	Uwagi
Dostarczenie detali	60	10	nieregularne
Sortowanie ręczne	120	5	operator
Kontrola wizyjna	90	7	błędy systemu
Robot pick&place	30	20	bardzo szybki
Pakowanie manualne	180	3	wąskie gardło
Transport końcowy	60	10	stabilny

- Wyznaczenie czasu cyklu całego procesu

Zasada: czas cyklu linii wyznacza najwolniejsze stanowisko. Najdłuższy czas operacji ma: pakowanie manualne = 180 sekund

Odpowiedź: czas cyklu całego procesu wynosi 180 sekund na sztukę.

Interpretacja: oznacza to, że cała linia może dostarczać jedną sztukę gotowego wyrobu co 180 sekund, niezależnie od tego, że część stanowisk pracuje znacznie szybciej.

- Obliczenie wydajności linii

Wydajność można obliczyć jako:

$$\text{Wydajność} = \frac{1}{180\ \text{s}} = 0{,}00556\ \text{szt./s}$$

Po przeliczeniu na minuty:

$$0{,}00556 \times 60 = 0{,}33\ \text{szt./min}$$

Po przeliczeniu na godzinę:

$$0{,}33 \times 60 = 20\ \text{szt./h}$$

Odpowiedź: wydajność linii wynosi około 0,33 szt./min, czyli 20 szt./h.

- Obliczenie wykorzystania robota

Robot pracuje z czasem operacji 30 s/szt., czyli jego zdolność wynosi:

$$\frac{60}{30} = 2 \text{ szt./min}$$

Linia osiąga rzeczywistą wydajność:

$$0{,}33 \text{ szt./min}$$

Wykorzystanie robota:

$$\frac{0{,}33}{2} \times 100\% = 16{,}5\%$$

Odpowiedź: robot jest wykorzystywany tylko w około 16,5%.

Interpretacja: to bardzo niski poziom wykorzystania. Pokazuje, że potencjał automatyzacji nie został dopasowany do reszty procesu.

- Identyfikacja problemów na podstawie danych

Wnioski:

1.Wąskie gardło: pakowanie manualne, ponieważ ma najdłuższy czas.

2. Nierównowaga linii: robot i inne stanowiska pracują szybciej niż pakowanie.

3. Tworzenie zapasów międzyoperacyjnych: detale będą się gromadziły przed pakowaniem.

4. Przestoje automatyki: robot będzie okresowo zatrzymywany.

5. Niski zwrot z inwestycji w automatyzację: skoro robot pracuje tylko przez część czasu, system nie jest zoptymalizowany.

6. Wariant 1: drugi operator pakowania

Zakładamy, że drugi operator skraca czas pakowania z 180 s do 90 s.

Nowy czas cyklu:

Najwolniejsze stanowisko po zmianie:

sortowanie ręczne = 120 spakowanie = 90 s
kontrola = 90 s

Wąskim gardłem staje się sortowanie ręczne = 120 s

Nowa wydajność linii:

$$\frac{1}{120} = 0{,}00833 \text{ szt./s}$$
$$0{,}00833 \times 60 = 0{,}5 \text{ szt./min}$$
$$0{,}5 \times 60 = 30 \text{ szt./h}$$

Odpowiedź: po wprowadzeniu drugiego operatora wydajność wzrasta z 20 szt./h do 30 szt./h, czyli o 50%. W poprzedniej wersji roboczej można było intuicyjnie założyć duży wzrost, ale po uwzględnieniu całej linii widać, że ograniczeniem staje się inny etap. To ważna lekcja: usprawnienie jednego stanowiska nie zawsze daje proporcjonalny wzrost wydajności całego procesu.

7. Wariant 2: robot pakujący

Zakładamy, że czas pakowania spada do 40 s.

Nowe czasy:

- dostarczenie: 60 s
- sortowanie: 120 s
- kontrola: 90 s
- robot pick-and-place: 30 s
- pakowanie: 40 s
- transport: 60 s

Nowe wąskie gardło: Sortowanie ręczne = 120 s

Nowa wydajność:

Tak jak wyżej:

$$\frac{1}{120} = 0{,}5 \text{ szt./min}$$

czyli

30 szt./h

Odpowiedź: wprowadzenie robota pakującego bez zmiany sortowania zwiększa wydajność linii do 30 szt./h. To pokazuje, że kosztowna inwestycja nie daje pełnego efektu, jeśli nie zostanie poprzedzona analizą całego procesu.

8. Wariant 3: usprawnienie systemowe

Zakładamy:

- usprawnienie sortowania: 120 s → 70 s
- poprawa kontroli wizyjnej: 90 s → 60 s
- pakowanie zautomatyzowane: 40 s
- synchronizacja stanowisk i ograniczenie zatrzymań

Nowe czasy:

- dostarczenie: 60 s
- sortowanie: 70 s
- kontrola: 60 s
- robot: 30 s
- pakowanie: 40 s
- transport: 60 s

Nowe wąskie gardło: sortowanie = 70 s

Nowa wydajność:

$$\frac{1}{70} = 0{,}01429 \text{ szt./s}$$

$$0{,}01429 \times 60 = 0{,}86 \text{ szt./min}$$

$$0{,}86 \times 60 \approx 51 \text{ szt./h}$$

Odpowiedź: po usprawnieniu systemowym wydajność wzrasta do około 51 szt./h. Najlepsze efekty daje nie pojedyncza inwestycja, lecz zintegrowane usprawnienie całego procesu.

Podsumowanie

Opis procesu: proces obejmuje dostarczenie komponentów, sortowanie, kontrolę wizyjną, pobór przez robota, pakowanie oraz transport końcowy. Linia ma charakter częściowo zautomatyzowany.

Identyfikacja problemów: najważniejszym problemem jest wąskie gardło na stanowisku pakowania. Dodatkowo występują błędy systemu wizyjnego, przestoje robota i zapasy międzyoperacyjne.

Analiza danych: najdłuższy czas operacji wynosi 180 s, dlatego wydajność linii to 20 szt./h. Robot ma bardzo niski poziom wykorzystania.

Propozycje usprawnień: zautomatyzowanie pakowania, synchronizacja systemu sterowania, poprawa kontroli wizyjnej oraz reorganizacja stanowisk.

Najlepszy wariant: najbardziej efektywny jest wariant systemowy, ponieważ nie tylko eliminuje jedno ograniczenie, ale poprawia cały proces.

Uzasadnienie: podejście procesowe pokazuje, że optymalizacja pojedynczego stanowiska nie wystarcza. Skuteczne usprawnienie wymaga analizy przepływu całego strumienia wartości

Bibliografia

Davenport, T. H. (1993): Process innovation: Reengineering work through information technology. Harvard Business School Press.

Goldratt, E. M. (1990):What is this thing called theory of constraints and how should it be implemented? North River Press.

Hammer, M., & Champy, J. (1993): Reengineering the corporation: A manifesto for business revolution. Harper Business.

Hirano, H. (1995):5 pillars of the visual workplace: The sourcebook for 5S implementation. Productivity Press.

Imai, M. (1986): Kaizen: The key to Japan's competitive success. McGraw-Hill.

Ishikawa, K. (1985):What is total quality control? The Japanese way. Prentice Hall.

Juran, J. M. (1988): Juran on planning for quality. Free Press.

Kubiński, P. (red.). (2017): Innowacje w biznesie – nowe modele i nowe praktyki. Exante.

Liker, J. K. (2004): The Toyota way: 14 management principles from the world's greatest manufacturer. McGraw-Hill.

Miles, L. D. (1961): Techniques of value analysis and engineering. McGraw-Hill.

Matusiak, K. B. (Red.). (2006): Innowacje i przedsiębiorczość dla przyszłości. SOOIPP.

Ohno, T. (1988). Toyota production system: Beyond large-scale production. Productivity Press.

OECD. (2005): Oslo manual: Guidelines for collecting and interpreting innovation data (3rd ed.). OECD Publishing.

Rother, M., Shook, J. (2003): Learning to see: Value stream mapping to add value and eliminate muda (version 1.4). Lean Enterprise Institute.

Schumpeter, J. A. (1960): Teoria rozwoju gospodarczego. Państwowe Wydawnictwo Naukowe.

Trzepizur, P. (2016): Zarządzanie innowacjami w małych i średnich przedsiębiorstwach. Zeszyty Naukowe Politechniki Częstochowskiej. Zarządzanie, 24(1), 52–62.

Wiśniewska, J., Janasz, K. (Red.). (2023): Innowacje i kreatywność we współczesnym przedsiębiorstwie. Wydawnictwo IVG.

Womack, J. P., Jones, D. T. (2001): Lean thinking: Banish waste and create wealth in your corporation (2nd ed.). Free Press.