



Integrated design of safety and control systems for nonlinear safety-critical processes[☆]

Xiaodong Cui^a, Arthur Khodaverdian^a, Panagiotis D. Christofides^{a,b,*}

^a Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095, USA

^b Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095-1592, USA

ARTICLE INFO

Keywords:

Process operational safety
Model predictive control
Safety systems
Nonlinear systems
Safety-critical processes

ABSTRACT

Safe operation of nonlinear safety-critical processes requires coordination of feedback control, protection against abnormally large disturbances using safety systems, shutdown, and recovery after repair. This work develops an integrated control and safety system architecture for these operating stages. For nominal operation, the architecture uses Safeness Index-informed Lyapunov-based model predictive control (MPC); following supervisory activation, the supervisor coordinates measurement-based active protection, shutdown, and recovery after repair. Under the stated assumptions, the analysis demonstrates closed-loop stability during nominal operation, preservation of physical safety during certified nonshutdown operation, and finite-time entry into the nominal region after repair. In a methyl isocyanate hydrolysis reactor case study, feedback cooling alone failed to contain the tested abnormal disturbance, whereas both conventional on/off and smart safety systems maintained containment. Relative to conventional on/off operation, the smart policy reduced the mean and final values of the normalized discharge metric by 34.55% and 40.78%, respectively, without increasing the observed maximum reactor temperature. Nominal handoff occurred 272 s after repair. These results demonstrate that physical safety, reduced material loss, and recovery can be coordinated within a unified architecture.

1. Introduction

Safe operation is a fundamental requirement for chemical processes, which must remain within admissible operating envelopes defined by critical variables rather than be evaluated only for nominal steady-state performance (Stauffer and Chastain-Knight, 2021; Senthilvelan et al., 2025; Albalawi et al., 2018). These admissible operating envelopes are commonly expressed through Safe Operating Limits (SOLs), which specify allowable ranges for critical process parameters such as temperature, pressure, level, flow, pH, and composition (Stauffer and Chastain-Knight, 2021; Senthilvelan et al., 2025). For processes subject to the Occupational Safety and Health Administration (OSHA) Process Safety Management (PSM) standard, process safety information is required to include safe upper and lower limits for variables such as temperature, pressure, flow, and composition, together with an evaluation of the consequences of deviations from these limits (Occupational Safety and Health Administration, 1992). When these limits are exceeded, operation should transition from ordinary troubleshooting to predetermined corrective or protective actions that return the process system to a safe state (Stauffer and Chastain-Knight, 2021; Senthilvelan

et al., 2025); therefore, safe operation should be treated as a continuous operating objective during both normal and abnormal process operation, especially when disturbances, equipment faults, or actuator and sensor limitations drive process variables toward the boundaries of the admissible operating envelope (Albalawi et al., 2018; Wang et al., 2023).

Recent chemical process incidents demonstrate that temperature excursions, pressure buildup, loss of containment, and failures of protective systems continue to present significant operational safety challenges, as summarized in Table 1 (U.S. Chemical Safety and Hazard Investigation Board, 2026a, 2025, 2026b, 2023a,b). The Givaudan and Optima Belle incidents emphasize the need to characterize reactive hazards so that process equipment and relief systems account for credible decomposition and overpressure scenarios (U.S. Chemical Safety and Hazard Investigation Board, 2026a, 2023a). The Marathon and Intercontinental Terminals Company incidents demonstrate the importance of connecting appropriate measurements and operating limits to alarms, interlocks, and predetermined remote protective actions; timely isolation can also limit accident escalation and the quantity of material released (U.S. Chemical Safety and Hazard Investigation Board, 2025,

[☆] This article is part of a Special issue entitled: 'Festschrift Y. Arkun' published in Chemical Engineering Research and Design.

* Corresponding author at: Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095, USA.

E-mail address: pdc@seas.ucla.edu (P.D. Christofides).

Table 1
Representative chemical process incidents and associated safety issues.

Incident	Main safety issue	Relevant safeguard
Givaudan Sense Colour, 2024 (U.S. Chemical Safety and Hazard Investigation Board, 2026a)	Runaway decomposition and reactor rupture after the relief capacity was exceeded.	Reactive hazard assessment and adequate relief capacity.
Marathon Martinez Renewable Fuels, 2023 (U.S. Chemical Safety and Hazard Investigation Board, 2025)	Fired-heater tube overheating and rupture following a loss of process flow.	Flow measurement, alarms, safety interlocks, and remote shutdown.
Dow Plaquemine Glycol II Unit, 2023 (U.S. Chemical Safety and Hazard Investigation Board, 2026b)	Ignition propagated through pressure relief piping into process equipment.	Monitoring of the inert atmosphere and appropriate relief system design.
Optima Belle, 2020 (U.S. Chemical Safety and Hazard Investigation Board, 2023a)	Uncharacterized thermal decomposition and dryer overpressure.	Process knowledge and thermal hazard assessment.
Intercontinental Terminals Company, 2019 (U.S. Chemical Safety and Hazard Investigation Board, 2023b)	An undetected flammable release and inability to isolate tank inventories.	Release detection and remote emergency isolation.

2023b). The Dow incident further illustrates that auxiliary protection equipment can itself become an escalation pathway when its operating condition is not adequately maintained or monitored (U.S. Chemical Safety and Hazard Investigation Board, 2026b). Although the initiating mechanisms differ, these incidents reveal a common sequence: a hazardous deviation develops, the condition must be recognized promptly, and the response must shift from ordinary process control to dedicated protective action before escalation. This common need motivates the integration of operational safety with feedback control, where the controller maintains desirable process operation when sufficient control authority is available, while a separate safety system monitors the process safety condition and initiates appropriate protective actions when feedback control is no longer sufficient (Albalawi et al., 2018; Zhang et al., 2018a; Leveson and Stephanopoulos, 2014).

Model Predictive Control (MPC) has become a widely used advanced control framework for chemical process systems because it predicts future process behavior using a dynamic model and computes control actions by solving a finite-horizon constrained optimization problem (Qin and Badgwell, 2003; Christofides et al., 2013; Ellis et al., 2014). A key advantage of MPC is that operational requirements can be incorporated directly as constraints on manipulated inputs, process states, output variables, or input-rate changes, which allows the controller to account for multivariable interactions while maintaining operation within admissible limits (Qin and Badgwell, 2003; Christofides et al., 2013). For example, in protonic membrane steam methane reforming systems, model-based control including MPC has been used to coordinate thermal hydrogen generation, electrochemical hydrogen recovery, and steam generation, while safety-relevant restrictions such as extraction-current or current-density limits can be incorporated to avoid undesirable electrochemical transients and protect thermo-electrochemical performance (Cui et al., 2024; Peters et al., 2026). In another example, Nonlinear Model Predictive Control (NMPC) has been applied to exothermic semi-batch reactors by incorporating runaway criteria into the control problem, allowing the reagent feed flow rate to be manipulated to avoid thermal runaway (Kummer et al., 2020). Similarly, Hamilton–Jacobi reachability has been combined with MPC to identify avoidable sets associated with thermal runaway and impose these unsafe regions as constraints, thereby excluding trajectories that would lead to excessive reactor temperature or unsafe operation (Kanavalau and Lall, 2022). Beyond conventional constrained MPC, more explicit safety-oriented formulations have been developed, including Safeness Index-based Model Predictive Control (SI-MPC), Control Barrier Function-based Model Predictive Control (CBF-MPC), and Control Lyapunov-Barrier Function-based Model Predictive Control (CLBF-MPC), where safety metrics, barrier conditions, or safe-region

constraints are embedded into the optimization problem (Albalawi et al., 2017, 2018; Wu et al., 2019).

Control Barrier Function (CBF) methods also provide a direct means of enforcing safety through constraints on admissible control inputs. Robust formulations for sampled-data systems account for bounded disturbances, measurement errors, and state evolution between sampling instants (Oruganti et al., 2024). Online disturbance estimation has also been incorporated into CBF formulations, where estimates of unknown contributions to the barrier dynamics and their associated error bounds are used to construct robust safety constraints (Daş and Murray, 2022).

More recently, Reinforcement Learning (RL)-based controllers have been investigated to reduce online computational effort through offline training, while safety-filtered or gated RL architectures use barrier functions, backup controllers, or supervisory filters to reject unsafe actions and convert unconstrained learned policies into constrained or safe RL policies (Cheng et al., 2019; Wang and Wu, 2024; Cui et al., 2026a,b). Despite these advances, MPC- and RL-based safety controllers remain controller-side safety mechanisms. Industrial considerations such as protective-action selection, material loss, emergency shutdown cost, interactions with alarms and interlocks, and recovery after the abnormal condition is resolved can, in principle, be incorporated into an MPC formulation; however, the resulting safety guarantee still depends on model accuracy and optimization feasibility for MPC, representative training data for RL, and reliable measurements and sufficient feedback actuator authority for both. Moreover, these methods primarily focus on constraint satisfaction, safe-set invariance, stabilization, or economic optimization under safety constraints. Therefore, integrating control systems and safety systems provides a complementary architecture in which the controller maintains desirable operation when feasible, while the safety system provides an additional protection layer when the model, optimization, or actuator conditions required for controller-based safety are no longer available (Zhang et al., 2018a; Albalawi et al., 2018).

Several studies have begun to investigate how feedback controllers and safety systems can be coordinated to improve process operational safety. For example, model predictive safety systems have been proposed to proactively detect operational hazards and support alarm triggering before the process reaches unsafe conditions (Ahooyi et al., 2016). The dynamic interaction between feedback control and safety systems has also been analyzed for chemical process applications, including a methyl isocyanate (MIC) hydrolysis reactor where a Lyapunov-based Model Predictive Control (LMPC) system uses integration with safety-system activation to avoid thermal runaway, and a high-pressure flash drum where controller tuning was modified according to the activation status of a pressure-relief safety system

(Zhang et al., 2018a,b). In addition, SI-MPC has been integrated with safety relief valve activation to coordinate controller behavior with discrete protective actions during unsafe operation (Zhang et al., 2019). These studies demonstrate that considering the interaction between the control layer and the safety layer can improve closed-loop operational safety compared with treating them as completely independent components. Within such an architecture, different admissible protective intensities can result in different material losses and intervention durations. This motivates adjusting safety-system action according to measured process behavior while keeping performance adjustment subject to explicit safety requirements. The coordination must also extend to shutdown requests and recovery after repair, including the conditions for resuming nominal control.

To address this coordination problem, this work develops an integrated control and safety system architecture for nonlinear safety-critical processes. During normal operation, Safeness Index-informed Lyapunov-based MPC (SI-LMPC) pursues the nominal control objective while penalizing predicted movement toward less safe operating conditions. Following supervisory activation, the active safety policy does not propagate the nominal process model to predict the abnormal dynamics. Instead, sampled measurements are combined with a certified characterization of the incremental effect of the safety input that is independent of the disturbance. Together, they determine a certified lower bound on the protective action intensity required to satisfy the prescribed bound on the safety-certificate rate over the next sampling interval whenever a feasible intensity exists. Completed safety evaluations update a separate candidate action intensity, and the applied intensity is selected as the larger of this candidate and the current certified lower bound. The framework also distinguishes active safety operation from recovery after repair. After repair has been confirmed and the disturbance again satisfies the bound assumed for normal operation, a recovery policy first attempts recovery using feedback control alone and supplements it with the safety input only when necessary. Nominal MPC is resumed only after the process has entered the nominal Lyapunov region (closed-loop stability region under MPC) and the handoff conditions are satisfied. Under the stated assumptions, the analysis establishes (i) forward invariance of the nominal Lyapunov region and ultimate boundedness under normal-operation disturbances; (ii) preservation of the physical safe set during active safety operation, provided that the abnormal disturbance remains within the certified envelope and the supervisor can apply a certified safety action without invoking shutdown; and (iii) entry into the nominal Lyapunov region within finite time under the stated recovery assumptions. A strongly exothermic MIC hydrolysis process susceptible to thermal runaway is used to illustrate the methodology.

2. Preliminaries

2.1. Notation

The actual state and the nominal predicted state are denoted by x and $\tilde{x}$, respectively. The transpose and Euclidean norm of a vector x , the set of real numbers, and set difference are denoted by $x^\top$, $\|x\|$, $\mathbb{R}$, and $\Omega_1 \setminus \Omega_2$, respectively. For a set Ω , $\text{int}(\Omega)$ denotes its interior. A generic function is denoted by $f(\cdot)$. The sampling instants are $t_k = t_0 + k\Delta$, $k \in \mathbb{Z}_{\geq 0}$, where t_0 is the initial sampling instant and $\Delta > 0$ is the sampling period. The set of input functions that are constant over each sampling interval is denoted by $S(\Delta)$. For a given state $x(t_k)$ and input trajectory $u(\cdot)$, $\tilde{x}(t | t_k)$ denotes the nominal predicted state trajectory generated from $x(t_k)$ under $u(\cdot)$. A function $\beta : [0, a) \rightarrow [0, \infty)$, where $a > 0$, is said to be of class- $\mathcal{K}$ if it is continuous, strictly increasing, and satisfies $\beta(0) = 0$. For $\beta \in \mathcal{K}$, β^{-1} denotes its inverse on its range.

2.2. Class of systems

This work considers safety-critical nonlinear processes equipped with a feedback control system and a separately actuated safety system. During normal operation, the process dynamics are represented by

$$\dot{x}(t) = F(x(t), u(t), u_s(t), w(t)), \quad (1)$$

where $x(t) \in D \subseteq \mathbb{R}^n$ is the process state, $u(t) \in U \subset \mathbb{R}^m$ is the feedback control input, and $u_s(t) \in U_s \subset \mathbb{R}^p$ is the safety-system input. The domain and admissible-input sets satisfy $0 \in D$, $0 \in U$, and $0 \in U_s$. The admissible input sets U and U_s are bounded and represent the physical operating limits of the feedback control and safety actuators, respectively. The variables x and u are defined in deviation form with respect to the nominal equilibrium state and input, respectively, whereas u_s is expressed relative to its inactive value, so that $u_s = 0$ when the safety actuator is not in use.

The signal $w(t) \in \mathbb{R}^r$ represents model mismatch, exogenous disturbances, and the equivalent effects of faults not explicitly included in the process model. During normal operation, it satisfies

$$w(t) \in \mathcal{W}_N := \{w \in \mathbb{R}^r \mid \|w\| \leq W_{\max}\}, \quad (2)$$

where $W_{\max} > 0$ is the largest disturbance magnitude under nominal operation. An abnormal event is characterized by $w(t) \notin \mathcal{W}_N$, or equivalently $\|w(t)\| > W_{\max}$.

The process state and any additional safety-relevant measurements are available at the sampling instants t_k . The feedback control input is implemented in sample-and-hold fashion, such that $u(t) = u_k$ for $t \in [t_k, t_{k+1})$. The resulting physical safety-system input (u_s) may depend on the current state, which may vary continuously with time, and thus may not remain constant within a sampling interval. The vector field F is locally Lipschitz in x over the normal operating domain, and the process trajectory is assumed to remain well defined over the interval under consideration.

The nominal model with the safety system inactive is denoted by

$$F^0(x, u) := F(x, u, 0, 0). \quad (3)$$

The system is written in deviation-variable form such that $F^0(0, 0) = 0$. The model F^0 is used for nominal controller design and prediction.

Remark 1. The robust forward-invariance guarantee developed in Section 3.4 for the physical process under nominal MPC relies on $w(t) \in \mathcal{W}_N$ and on the model and Lyapunov bounds certified over the nominal operating region. During an abnormal event, these bounds need not remain valid, so the nominal model may not represent the physical dynamics accurately enough for that guarantee to apply. If the event drives the state outside the region on which the nominal Lyapunov conditions are certified, no nominal decrease or forward-invariance property is claimed there. This limitation motivates the separate safety system, which uses measurements and a certified characterization of the safety-input effect rather than prediction of the abnormal dynamics.

2.3. Process safety characterization

Let $\mathcal{X}_{\text{safe}}^{\text{phys}} \subset D$ denote the closed set of states satisfying the physical safety requirements of the process. Safe operation requires $x(t) \in \mathcal{X}_{\text{safe}}^{\text{phys}}$ throughout the operating interval.

To characterize this additional information, this work adopts the Safeness Index concept introduced in Albalawi et al. (2017). Let I_k denote the safety-relevant information causally available at t_k . Depending on the application, I_k may include the current state, additional measurements, a finite measurement history, changes in selected variables, estimated quantities, or diagnostic information. The Safeness Index is defined by

$$S_k = S(I_k) \in [-1, 1]. \quad (4)$$

The value $S_k = -1$ represents the most favorable condition, whereas $S_k = 1$ represents the supervisory limiting condition at which continued operation is not permitted. Larger intermediate values indicate a smaller safety margin.

The physical safe set $\mathcal{X}_{\text{safe}}^{\text{phys}}$ specifies the state constraint that must be satisfied during process operation, whereas the Safeness Index S_k is a metric that quantifies the proximity to unsafe operating conditions based on observed process information (I_k). Because I_k may contain measurement history or other information that is not uniquely determined by $x(t_k)$, the same state may correspond to different Safeness Index values. Consequently, Safeness Index levels do not, in general, define fixed state-space boundaries and are not interchangeable with the boundary of $\mathcal{X}_{\text{safe}}^{\text{phys}}$. Any specific relationship between S_k and membership in $\mathcal{X}_{\text{safe}}^{\text{phys}}$ must be established by the process-specific construction of the Safeness Index.

Remark 2. Like the physical safe set $\mathcal{X}_{\text{safe}}^{\text{phys}}$, the construction of S is process dependent and may be based on physical knowledge, formal safety analyses, operating data, and validated measurements. This subsection establishes only its general interpretation. The thresholds and the actions associated with them are introduced later as part of the safety-system design.

2.4. Stabilizability assumption

The following assumption states the nominal stabilizability requirements for the disturbance-free, safety-inactive model $\dot{x} = F^0(x, u)$ defined in Eq. (3). It jointly requires the existence of a feedback law and an associated Lyapunov function satisfying the stated conditions on Ω_ρ . These conditions ensure that each nonlinear process system considered admits a stabilizing feedback controller and provides a Lyapunov characterization of a certified closed-loop stability region.

Assumption 1. There exist a locally Lipschitz feedback law $\Phi : D \rightarrow U$, with $\Phi(0) = 0$, a continuously differentiable Lyapunov function $V : D \rightarrow \mathbb{R}_{\geq 0}$, a constant $\rho > 0$, and class- $\mathcal{K}$ functions $\beta_1, \beta_2, \beta_3$, and β_4 such that the following inequalities hold for every $x \in D$,

$$\beta_1(\|x\|) \leq V(x) \leq \beta_2(\|x\|), \quad (5a)$$

$$\frac{\partial V(x)}{\partial x} F^0(x, \Phi(x)) \leq -\beta_3(\|x\|), \quad (5b)$$

$$\left\| \frac{\partial V(x)}{\partial x} \right\| \leq \beta_4(\|x\|). \quad (5c)$$

In addition, we denote level sets of the Lyapunov function within this open region as $\Omega_\eta := \{x \in D \mid V(x) \leq \eta\}$ where η is an arbitrary constant. Then, for any $\rho > 0$, there exist positive constants M_F, L_x, L_w, L'_x , and L'_w such that, for every $x, x' \in \Omega_\rho, u \in U$, and $w \in \mathcal{W}_N$,

$$\|F(x, u, 0, w)\| \leq M_F, \quad (6a)$$

$$\|F(x, u, 0, w) - F^0(x', u)\| \leq L_x \|x - x'\| + L_w \|w\|, \quad (6b)$$

$$\left\| \frac{\partial V(x)}{\partial x} F(x, u, 0, w) - \frac{\partial V(x')}{\partial x'} F^0(x', u) \right\| \leq L'_x \|x - x'\| + L'_w \|w\|. \quad (6c)$$

Under **Assumption 1**, Ω_ρ is forward invariant for the nominal closed-loop system $\dot{x} = F^0(x, \Phi(x))$, and every trajectory initialized in this set converges asymptotically to the origin. Its inclusion in $\text{int}(\mathcal{X}_{\text{safe}}^{\text{phys}})$ provides a physical safety margin for nominal operation.

Remark 3. For the actual process, $w(t) \in \mathcal{W}_N$ is allowed. In that case, the bounds in Eq. (6), together with the sampled-data conditions stated in part 1 of **Theorem 1**, establish robust forward invariance of Ω_{ρ^*} and ultimate boundedness under normal-operation disturbances.

2.5. Model predictive control

Model predictive control is a widely used optimization-based framework for computing control actions for constrained dynamical systems. At each sampling instant t_k , the MPC computes a piecewise-constant input trajectory $u(\cdot) \in S(\Delta)$ over a prediction horizon of N sampling periods, where $N \in \mathbb{Z}_{>0}$, by solving the following optimal control problem:

$$\min_{u(\cdot) \in S(\Delta)} \int_{t_k}^{t_k + N\Delta} L(\tilde{x}(t \mid t_k), u(t)) dt \quad (7a)$$

$$\text{s.t. } \dot{\tilde{x}}(t \mid t_k) = F^0(\tilde{x}(t \mid t_k), u(t)) \quad (7b)$$

$$\tilde{x}(t_k) = x(t_k) \quad (7c)$$

$$\tilde{x}(t \mid t_k) \in \mathcal{X}_c, \quad \forall t \in [t_k, t_k + N\Delta] \quad (7d)$$

$$u(t) \in \mathcal{U}_c, \quad \forall t \in [t_k, t_k + N\Delta]. \quad (7e)$$

In Eq. (7a), $L(x, u)$ denotes the stage cost, which is selected according to the control objective, such as regulation, reference tracking, or economic performance. The predicted state trajectory $\tilde{x}(t \mid t_k)$ is generated by the nominal model in Eq. (7b) from the measured state $x(t_k)$ in Eq. (7c), subject to the state and input constraints in Eqs. (7d) and (7e), where $\mathcal{X}_c \subseteq D$ and $\mathcal{U}_c \subseteq U$ denote the corresponding constraint sets. After Eq. (7) is solved, only the first control action of the optimal input trajectory is applied to the process and held constant over $[t_k, t_{k+1})$. At t_{k+1} , the prediction horizon is shifted to $[t_{k+1}, t_{k+1} + N\Delta]$, and the problem is solved again using the newly measured state. This repeated optimization and first-action implementation is the receding-horizon approach.

Remark 4. The prediction model in Eq. (7b) corresponds to the nominal, safety-inactive dynamics obtained by setting $u_s \equiv 0$ and $w \equiv 0$. It is therefore the same model with respect to which the reference controller Φ and the Lyapunov function V are defined in Section 2.4. The effects of nonzero safety-system inputs are considered separately in the integrated control-safety formulation.

3. Integration of safety system and feedback control

This section develops the policies for nominal control, active safety operation, shutdown, and recovery after repair, with their coordination and switching logic summarized in Section 3.4.

3.1. Nominal predictive control informed by the safeness index

Nominal operation should be permitted only while the state remains in a region for which the Lyapunov-based design is certified and the available safety information remains below the first supervisory threshold. Accordingly, define the inner Lyapunov level set $\Omega_{\rho^*} := \{x \in D \mid V(x) \leq \rho^*\}$ as the nominal Lyapunov region, the state-based supervisory boundary, and the recovery target. The set Ω_ρ introduced in **Assumption 1** characterizes the region in which the nominal stability conditions are certified. The level ρ^* is selected below ρ to define the practical boundary for nominal operation, taking into account physical safety limits, bounded excursions caused by normal-operation disturbances, and the desired margin for supervisory action. It is additionally required that $\Omega_{\rho^*} \subset \text{int}(\mathcal{X}_{\text{safe}}^{\text{phys}})$. A sufficient condition ensuring robustness of the selected nominal region is provided in part 1 of **Theorem 1**.

Select a first supervisory threshold $0 < S_{\text{th},1} < 1$. At t_k , nominal operation is permitted when the supervisory mode is inactive, $x(t_k) \in \Omega_{\rho^*}$, and $S_k < S_{\text{th},1}$. The safety-system input is then maintained at $u_s = 0$. Because $S_k = S(I_k)$ may depend on measurement history or other information that is not generated by the nominal model, the future measured Safeness Index sequence is not assumed to be predictable from F^0 . The controller therefore predicts only the nominal

state trajectory using F^0 and includes a nonnegative safety penalty $\mathcal{J}_S$ based on this trajectory and the information available at t_k . At each sampling instant eligible for nominal operation, it solves

$$\min_{u(\cdot) \in \mathcal{S}(\Delta)} \int_{t_k}^{t_k + N\Delta} L(\tilde{x}(t | t_k), u(t)) dt + \mathcal{J}_S(\tilde{x}(\cdot | t_k), u(\cdot); \mathcal{I}_k) \quad (8a)$$

$$\text{s.t. } \dot{\tilde{x}}(t | t_k) = F^0(\tilde{x}(t | t_k), u(t)) \quad (8b)$$

$$\tilde{x}(t_k | t_k) = x(t_k) \quad (8c)$$

$$\tilde{x}(t | t_k) \in \mathcal{X}_c, \quad t \in [t_k, t_k + N\Delta] \quad (8d)$$

$$u(t) \in \mathcal{U}_c, \quad t \in [t_k, t_k + N\Delta] \quad (8e)$$

$$\frac{\partial V(x(t_k))}{\partial x} F^0(x(t_k), u(t_k)) \leq \frac{\partial V(x(t_k))}{\partial x} F^0(x(t_k), \Phi(x(t_k))). \quad (8f)$$

The stage cost and constraint sets have the meanings introduced in Section 2.5. They are selected so that $\Omega_{\rho^*} \subseteq \mathcal{X}_c$ and $\Phi(x) \in \mathcal{U}_c$ for every $x \in \Omega_{\rho^*}$. The functional $\mathcal{J}_S$ assigns a larger cost when the nominally predicted state trajectory moves toward less favorable operating conditions. It is calculated using smooth process-specific measures of risk based on $\tilde{x}(\cdot | t_k)$, $u(\cdot)$, and the information in $\mathcal{I}_k$ available at t_k . These measures are used only in the control objective and are not predictions of the future measured Safeness Index sequence.

The constraint in Eq. (8f) requires the nominal Lyapunov time derivative generated by the first optimized input to be no greater than that generated by the reference controller Φ . We assume that Eq. (8) is feasible, attains a finite minimizer, and can be solved within each sampling period whenever nominal operation is permitted. Let $u^*(\cdot | t_k)$ denote an optimal input trajectory computed using the nominal prediction model F^0 . The receding-horizon implementation applies its first value, $u(t) = u^*(t_k | t_k)$, together with $u_s(t) = 0$, to the actual process over $[t_k, t_{k+1})$. Although Eq. (8b) is disturbance-free, the actual process may be subject to $w(t) \in \mathcal{W}_N$. The sufficient sampling and robustness conditions accounting for this disturbance are stated in part 1 of Theorem 1.

Additionally, the process may experience an abnormal event in which the disturbance exceeds these bounds, potentially resulting in deviation from Ω_{ρ^*} , in which case the supervisory mode is activated. More generally, the supervisory mode is activated when either of the following conditions is satisfied:

$$V(x(t_k)) > \rho^* \quad \text{or} \quad S_k \geq S_{\text{th},1}. \quad (9)$$

The resulting supervisory actions are introduced in the following subsection.

Remark 5. The disturbance classification is distinct from the supervisory mode. The condition $w(t) \in \mathcal{W}_N$ defines the normal disturbance regime, whereas the supervisory activation condition is an observable mode-selection rule. In particular, the Safeness Index may activate the supervisor while $x(t_k) \in \Omega_{\rho^*}$, allowing an early or conservative response before the state-based boundary is crossed. The state-based condition $V(x(t_k)) > \rho^*$ is treated as an observable indication that nominal operation may have been compromised and therefore prompts investigation of a possible abnormal condition.

3.2. Smart safety system design

Let t_{k_a} be the first sampling instant at which Eq. (9) is satisfied. With $S_{\text{th},1}$ defined in the preceding subsection, select two additional thresholds such that $0 < S_{\text{th},1} < S_{\text{th},2} < S_{\text{th},3} < 1$. Once activated, the supervisory mode remains latched until the abnormal condition has been resolved and the handoff conditions introduced later have been satisfied.

Physical containment after supervisory activation requires a physical safety metric, a protective feedback law, and a function that determines the physical safety-system input from the process state and

the selected action intensity. The required properties of these objects are collected in the following assumption.

Assumption 2. There exists a continuously differentiable physical safety certificate $c : D \rightarrow \mathbb{R}$, whose value is available from measurements at the sampling instants. Define

$$\mathcal{X}_A := \{x \in D \mid c(x) \leq 0\}, \quad \Omega_{\rho^*} \subset \text{int}(\mathcal{X}_A), \quad \mathcal{X}_A \subset \text{int}(\mathcal{X}_{\text{safe}}^{\text{phys}}). \quad (10)$$

The set $\mathcal{X}_A$ is compact. There also exist an admissible protective feedback law $u^{\text{safe}} : \mathcal{X}_A \rightarrow U$, a function $\kappa_s : D \times [0, 1] \rightarrow U_s$, and a maximum safety-system input $\bar{u}_s : D \rightarrow U_s$ such that, for every $x \in D$, $\kappa_s(x, 0) = 0$ and $\kappa_s(x, 1) = \bar{u}_s(x)$.

The certificate c defines the set in which physical containment is certified during active safety operation. The formulation below is stated for a single physical safety certificate. Unlike the nominal reference law Φ , which establishes nominal stabilization within Ω_{ρ^*} , u^{safe} is applied after supervisory activation to support containment in $\mathcal{X}_A$.

For each state x and normalized safety-action intensity $\alpha \in [0, 1]$, $\kappa_s(x, \alpha)$ gives the corresponding physical safety-system input. At each sampling instant t_k , the supervisor selects an intensity α_k , which is held over $[t_k, t_{k+1})$. The resulting physical safety-system input is $u_s(t) = \kappa_s(x(t), \alpha_k)$ and may therefore vary with the state within the sampling interval even though α_k remains fixed.

Assumption 2 collects conditions that can be checked during process-specific safety-system design. In practice, the certificate c may be constructed from validated safety limits on measured variables such as temperature, pressure, or concentration. A compact set $\mathcal{X}_A$ can be selected by restricting the certification to a bounded operating domain. The two inclusions in Eq. (10) provide margins between nominal operation, active safety operation, and the physical safety boundary. The protective feedback law and the relation $u_s = \kappa_s(x, \alpha)$ can be determined from admissible actuator operation, physical relations, equipment data, or offline testing. Thus, this assumption does not require prediction of the abnormal disturbance. It only requires the components used by the active-safety policy to be available and certified over the stated domain.

The following assumption specifies how the resulting physical safety-system input affects the rate of the safety certificate.

Assumption 3. During any sampling interval used to construct a rate estimate, the rate of the physical safety certificate can be decomposed, for almost every t , as

$$\frac{d}{dt} c(x(t)) = d^0(t) + g_s(x(t), u_s(t)), \quad (11)$$

where $d^0(t)$ collects all rate contributions other than the incremental effect of the safety-system input, including the unknown abnormal-process contribution and the contribution of the feedback control input applied over the interval. The incremental effect $g_s : D \times U_s \rightarrow \mathbb{R}$ is known, independent of the unknown disturbance, and satisfies $g_s(x, 0) = 0$. For every fixed x , the composite map $\alpha \mapsto g_s(x, \kappa_s(x, \alpha))$ is continuous and, for every $0 \leq \alpha^{(1)} \leq \alpha^{(2)} \leq 1$, satisfies

$$g_s(x, \kappa_s(x, \alpha^{(2)})) \leq g_s(x, \kappa_s(x, \alpha^{(1)})) \leq 0. \quad (12)$$

This assumption requires only characterization of the incremental effect of the engineered safety actuator; it does not require an online model of the unknown abnormal contribution $d^0(t)$. Such actuator information can be obtained from physical relations, equipment data, or offline calibration. The monotonicity condition merely states that increasing the protective-action intensity cannot worsen the rate of the physical safety certificate. It is therefore a basic certifiability requirement: without a known incremental safety-input effect, no selected action intensity could be certified against the unknown abnormal contribution.

The characterization of the safety-input effect is combined with sampled measurements to estimate the unknown contribution $d^0(t)$. No online model of $d^0(t)$ is used. For $k \geq k_a + 1$, the certificate rate over the completed sampling interval $[t_{k-1}, t_k]$ is approximated by the slope of the line connecting the two endpoint certificate measurements. If $\alpha_{k-1} \in [0, 1]$ denotes the known safety-action intensity held over $[t_{k-1}, t_k]$, subtracting the known incremental safety-input effect evaluated at the beginning of the interval gives

$$\hat{d}_k^0 := \frac{c(x(t_k)) - c(x(t_{k-1}))}{\Delta} - g_s(x(t_{k-1}), \kappa_s(x(t_{k-1}), \alpha_{k-1})). \quad (13)$$

The resulting $\hat{d}_k^0$ is a causal finite-difference estimate of an interval-representative value of the rate contribution d^0 in Eq. (11). This construction applies for every $\alpha_{k-1} \in [0, 1]$, including $\alpha_{k-1} = 0$.

The first protective interval is initialized without using a rate estimate. If the abnormal condition remains unresolved and no shutdown guard holds at t_{k_a} , apply $u(t) = u^{\text{safe}}(x(t_{k_a}))$ and select $q_{k_a} = 1$, $\alpha_{k_a} = 1$ when $S_{k_a} \geq S_{\text{th},1}$; otherwise select $q_{k_a} = 0$, $\alpha_{k_a} = 0$. The latter case corresponds to activation through $V(x(t_{k_a})) > \rho^*$ alone. The following capture condition covers the transition into supervisory operation and this first protective interval.

Assumption 4. For every abnormal realization covered by the non-shutdown safety certification, the activation index satisfies $k_a \geq 1$, and the state remains in $\mathcal{X}_{\text{safe}}^{\text{phys}}$ from the physical onset of the abnormal condition to t_{k_a} . In addition, the trajectory over the sampling interval immediately preceding supervisory activation satisfies $x(t) \in \mathcal{X}_A$ for all $t \in [t_{k_a-1}, t_{k_a}]$, with $x(t_{k_a-1}) \in \Omega_{\rho^*}$ and $x(t_{k_a}) \in \mathcal{X}_A$. If the first protective interval is applied, the corresponding initialized inputs described above keep $x(t) \in \mathcal{X}_A$ for every $t \in [t_{k_a}, t_{k_a+1}]$.

After completion of the first protective interval, Eq. (13) evaluated at t_{k_a+1} provides the first estimate $\hat{d}_{k_a+1}^0$ used for active-safety certification. The initialized intensity α_{k_a} is known in both branches. Earlier measurements may be available, but the present initialization uses the interval over which the protective feedback has been applied. The following certified upper bound supports this estimate and those constructed at subsequent sampling instants.

Assumption 5. There exists a certified margin $\varepsilon_c > 0$ such that, for every $k \geq k_a + 1$ during unresolved nonshutdown active-safety operation and every constant intensity $\alpha \in [0, 1]$ applied over $[t_k, t_{k+1}]$, the estimate $\hat{d}_k^0$ constructed from the completed sampling interval $[t_{k-1}, t_k]$ provides the bound

$$\frac{d}{dt} c(x(t)) \leq \hat{d}_k^0 + g_s(x(t_k), \kappa_s(x(t_k), \alpha)) + \varepsilon_c, \quad \text{for almost every } t \in [t_k, t_{k+1}]. \quad (14)$$

The margin accounts for measurement error, finite differencing, and variation of the unmeasured contribution and the known incremental effect over the completed and upcoming sampling intervals. It also accounts for changes in the feedback input between these intervals, including a transition from recovery to active safety operation if repair confirmation is withdrawn or a new abnormal condition is detected while the supervisory mode remains latched. When the preceding interval belongs to recovery, Eq. (13) uses the actually applied recovery intensity $\alpha_{k-1}^{\text{rec}}$ in place of α_{k-1} .

This assumption is a basic robustness condition for using a causal rate estimate in interval-based safety certification. It does not require an exact estimate of $d^0(t)$, a prediction of future states, or a model of the abnormal disturbance. It requires only that the estimation error and the variation occurring within one sampling interval admit a conservative finite upper bound over the certified operating domain. This condition is reasonable when the relevant measurement errors and interval variations are bounded and the sampling period and margin ε_c are selected accordingly. Without such an upper bound, neither the

condition for $u_s = 0$ nor the prescribed active-safety rate condition below could be certified over the following sampling interval.

For every $k \geq k_a + 1$ at which the active-safety policy is evaluated, define the certified upper bound on the constraint rate as

$$\hat{r}_k(\alpha) := \hat{d}_k^0 + g_s(x(t_k), \kappa_s(x(t_k), \alpha)) + \varepsilon_c. \quad (15)$$

Let $\eta_c \geq 0$ be the prescribed minimum decrease rate of the certificate. The quantity $\underline{\alpha}_k := \min \{1, \inf \{\alpha \in [0, 1] \mid \hat{r}_k(\alpha) \leq -\eta_c\}\}$ is the saturated required intensity, where $\inf \emptyset := +\infty$. If the feasible-intensity set inside the infimum is nonempty, continuity and the monotonicity in Eq. (12) imply that the infimum is attained and every $\alpha \in [\underline{\alpha}_k, 1]$ satisfies $\hat{r}_k(\alpha) \leq -\eta_c \leq 0$. If the set is empty, the definition gives $\underline{\alpha}_k = 1$, so the applied intensity defined below saturates at the maximum safety action. In this case, the prescribed rate condition is not claimed, and interval containment is provided by [Assumption 6](#).

When the safety-system input is inactive, the protective feedback input $u = u^{\text{safe}}(x(t_k))$ is retained and $u_s = 0$. This zero-input choice, after initialization, is certified for the next sampling interval only when

$$c(x(t_k)) + \Delta \max \{0, \hat{r}_k(0)\} \leq 0. \quad (16)$$

This condition permits an increase in c provided that the available safety margin is sufficient to keep $c(x(t)) \leq 0$ throughout the next sampling interval; it does not require $\hat{r}_k(0) \leq 0$. When Eq. (16) holds, integration of Eq. (14) with $\alpha = 0$ gives, for every $\tau \in [0, \Delta]$,

$$c(x(t_k + \tau)) \leq c(x(t_k)) + \tau \hat{r}_k(0) \leq c(x(t_k)) + \Delta \max \{0, \hat{r}_k(0)\} \leq 0. \quad (17)$$

Let $q_k \in \{0, 1\}$ denote the safety-action status selected for $[t_k, t_{k+1}]$, and set $q_k = 0$ while the supervisory mode is inactive. Hence, $q_{k_a-1} = 0$, and q_{k_a} is initialized as described above. For every $k \geq k_a + 1$ for which the active-safety policy is evaluated, set

$$q_k = \begin{cases} 1, & q_{k-1} = 0 \text{ and } (S_k \geq S_{\text{th},1} \text{ or Eq. (16) fails}), \\ 0, & q_{k-1} = 1, \quad S_k < 0, \quad \hat{r}_k(0) \leq 0, \\ & \text{and Eq. (16) holds,} \\ q_{k-1}, & \text{otherwise.} \end{cases} \quad (18)$$

Thus, failure of Eq. (16) initiates a safety action when it is inactive and prevents release when it is active. Release additionally requires $S_k < 0$ and a nonpositive certified zero-input rate. Returning the safety input to zero does not clear the supervisory mode or confirm repair. Each zero-input decision after initialization guarantees only the next sampling interval through Eq. (16). The resulting protective excursions need not be periodic. While the abnormal condition remains unresolved, the objective of the active-safety policy is containment in $\mathcal{X}_A$, rather than convergence to the nominal equilibrium.

Let $t^* > 0$ be the maximum total duration of latched supervisory operation, measured from t_{k_a} and including recovery after repair. At every sampling instant while the mode remains latched, shutdown is requested before initialization, active-safety updates, recovery, or nominal handoff if $t_k - t_{k_a} \geq t^*$ or $S_k > S_{\text{th},3}$. If neither condition holds, confirmed repair leads to the recovery procedure in the following subsection, while an unresolved abnormal condition leads to the active-safety policy. Any independent plant shutdown guard, if present, lies outside the proposed supervisory logic and retains overriding authority.

Assumption 6. For every $k \geq k_a + 1$ during unresolved active-safety operation for which neither proposed shutdown condition holds, if $q_k = 1$ and no $\alpha \in [0, 1]$ satisfies $\hat{r}_k(\alpha) \leq -\eta_c$, then the maximum safety action satisfies

$$c(x(t_k)) + \Delta \max \{0, \hat{r}_k(1)\} \leq 0. \quad (19)$$

When shutdown is not requested, $q_{k-1} = q_k = 1$, and $S_k \leq S_{\text{th},2}$, select a performance command $\alpha_k^{\text{cmd}} \in [0, 1]$. The applied intensity is

$$\alpha_k = \max \{\underline{\alpha}_k, \alpha_k^{\text{cmd}}\}. \quad (20)$$

Therefore, performance adjustment cannot reduce the applied intensity below the saturated required intensity. In particular, if the feasible-intensity set is empty, then $\alpha_k = 1$ and $\alpha_k = 1$.

The performance command may be adjusted using completed measurements of the consequences associated with safety-system operation. Let $J_j \geq 0$ denote the performance value obtained from the j th completed evaluation. Depending on the application, J_j may represent material loss, discharge rate, duration of safety-system use, operating cost, or another relevant consequence. The definition of an evaluation interval, the construction of J_j , the set of candidate commands, and the corresponding update rule are process dependent.

Any adjustment of α_k^{cmd} uses only completed measurements and must satisfy $\alpha_k^{\text{cmd}} \in [0, 1]$. Routine enforcement of the required lower bound and the prescribed maximum-action intervals form part of the evaluated policy, and their consequences are included in its performance value. Evaluations interrupted by repair, recovery, shutdown, a material change in the abnormal condition, or an exceptional override that prevents a comparable evaluation are not used for a performance update. Regardless of the selected performance criterion or update rule, the applied intensity in the adjustable branch remains defined by Eq. (20). Consequently, performance adjustment cannot reduce the safety action below the saturated required intensity $\underline{\alpha}_k$. No convergence to a globally optimal performance command is claimed.

While the abnormal condition remains unresolved and shutdown is not requested, maximum action $\alpha_k = 1$ is applied during the first interval of every transition from $q_{k-1} = 0$ to $q_k = 1$, including initialization, and during sustained activation when $S_k > S_{\text{th},2}$. The adjustable action in Eq. (20) is applied when $q_{k-1} = q_k = 1$ and $S_k \leq S_{\text{th},2}$. If the feasible-intensity set is empty, this rule automatically saturates at $\alpha_k = 1$, for which $\kappa_s(x(t), 1) = \bar{u}_s(x(t))$. If $q_k = 0$, the protective feedback law is retained with $u_s = 0$. The selected feedback input and, in the applicable safety branches, the intensity α_k are held over $[t_k, t_{k+1})$. The resulting physical safety-system input may vary with $x(t)$. With prescribed admissible shutdown laws $u^{\text{SD}} : D \rightarrow U$ and $u_s^{\text{SD}} : D \rightarrow U_s$, the policy is

$$(u(t), u_s(t)) = \begin{cases} (u^{\text{SD}}(x(t_k)), u_s^{\text{SD}}(x(t_k))), & \text{shutdown branch,} \\ (u^{\text{safe}}(x(t_k)), \bar{u}_s(x(t_k))), & \text{maximum-action branch,} \\ (u^{\text{safe}}(x(t_k)), \kappa_s(x(t), \alpha_k)), & \text{adjustable branch,} \\ (u^{\text{safe}}(x(t_k)), 0), & q_k = 0, \end{cases} \quad t \in [t_k, t_{k+1}). \quad (21)$$

The branches are evaluated in the stated priority order.

For the initial protective interval, [Assumption 4](#) guarantees containment under the prescribed initialization and supplies $x(t_{k_a+1}) \in \mathcal{X}_A$. For every subsequent unresolved nonshutdown interval with $k \geq k_a + 1$, $q_k = 0$ implies that Eq. (16) holds, whether the action remains inactive or is released. Thus, Eq. (17) guarantees $c(x(t)) \leq 0$ over that interval.

If $q_k = 1$ and the feasible-intensity set is nonempty, both the adjustable branch and any prescribed maximum-action branch apply an intensity satisfying $\hat{r}_k(\alpha_k) \leq -\eta_c$. Hence, Eq. (14) gives $\frac{d}{dt}c(x(t)) \leq -\eta_c \leq 0$ for almost every $t \in [t_k, t_{k+1})$. If the feasible-intensity set is empty, every active branch applies $\alpha_k = 1$, and [Assumption 5](#) and Eq. (19) guarantee interval containment without requiring a nonpositive certificate rate.

Consequently, [Assumption 4](#) covers physical safety up to activation and throughout the first protective interval. Induction from $k = k_a + 1$ using the preceding cases keeps $x(t) \in \mathcal{X}_A \subset \mathcal{X}_{\text{safe}}^{\text{phys}}$ during the subsequent certified nonshutdown active-safety intervals. The Safeness Index thresholds determine the supervisory decisions but do not themselves constitute the continuous-time physical safety proof. If a certification assumption ceases to hold without being detected, no nonshutdown safety claim is made. Safety of the shutdown transient requires a separate process-specific analysis.

Remark 6. Measurement noise may affect the Safeness Index, the evaluation of the safety certificate, and the rate estimate used by the protective policy. In practical implementations, filtering or other signal processing methods can be used to reduce these effects. The safety margins in the rate bound and the protective decision rules can also be increased to account for measurement uncertainty and residual estimation errors. Filter settings and margins should be selected according to sensor accuracy and the required response speed, with any delay introduced by filtering taken into account. The resulting measurement and estimation errors should be accounted for in the bounds and supervisory conditions used for safety certification.

3.3. Recovery to the stability region

Recovery is considered only after the abnormal condition has been removed, isolated, or reduced sufficiently for the normal-operation disturbance description to become valid again. Let $t_{k_f} \geq t_{k_a}$ be the first sampling instant at which this condition is confirmed. It is assumed that $w(t) \in \mathcal{W}_N$ throughout the recovery interval. Recovery laws may therefore be designed or validated using the normal-operation uncertainty bounds; no model of the abnormal process is propagated before t_{k_f} . Repair confirmation does not clear the supervisory mode, because the process state and the available safety information may not yet satisfy the conditions for nominal operation.

Let $\mathcal{X}_{\text{rec}}$ be a compact, certified recovery domain selected to satisfy

$$\Omega_{\rho^*} \subseteq \mathcal{X}_{\text{rec}} \subset \text{int}(\mathcal{X}_{\text{safe}}^{\text{phys}}), \quad \mathcal{X}_A \subseteq \mathcal{X}_{\text{rec}}. \quad (22)$$

The inclusion $\mathcal{X}_A \subseteq \mathcal{X}_{\text{rec}}$ links the active-safety and recovery phases. Because the active-safety policy keeps every certified nonshutdown trajectory in $\mathcal{X}_A$ until repair is confirmed, the state at t_{k_f} belongs to the certified recovery domain regardless of the phase of the protective excursion at which repair occurs. This inclusion is a design and certification requirement, rather than a consequence of physical safety alone.

The recovery input is allocated at every recovery sampling instant. Let $\mu_R : \mathcal{X}_{\text{rec}} \rightarrow U$ denote the recovery feedback law, and let $\mu_{\alpha,R} : \mathcal{X}_{\text{rec}} \times U \rightarrow [0, 1]$ denote the subsequent safety-intensity law. The sampled recovery inputs are selected according to

$$u_{R,k} = \mu_R(x(t_k)), \quad (23a)$$

$$\alpha_k^{\text{rec}} = \mu_{\alpha,R}(x(t_k), u_{R,k}), \quad (23b)$$

$$u(t) = u_{R,k}, \quad u_s(t) = \kappa_s(x(t), \alpha_k^{\text{rec}}), \quad t \in [t_k, t_{k+1}). \quad (23c)$$

The feedback input $u_{R,k}$ and safety intensity α_k^{rec} are held over $[t_k, t_{k+1})$, while the physical safety input varies with the state through κ_s . The feedback action $u_{R,k}$ is fixed first according to the recovery criterion. With this action fixed, $\mu_{\alpha,R}$ selects only the additional safety intensity required by the process-specific recovery certificate. Thus, $\alpha_k^{\text{rec}} = 0$ whenever the selected feedback action alone is sufficient over the current interval, whereas $\alpha_k^{\text{rec}} > 0$ makes the safety-system input supplement that action. Both quantities are recalculated at the next sampling instant, so a zero intensity does not define a separate recovery mode and does not prevent later use of u_s . The ordering in Eq. (23) describes the allocation of the two inputs and does not claim a global joint optimum over all $u \in U$ and $\alpha \in [0, 1]$.

The maps μ_R and $\mu_{\alpha,R}$ may be obtained from an offline feedback law or a certified online calculation. They are certified on $\mathcal{X}_{\text{rec}}$ only if, for every initial state in this set and every disturbance signal satisfying $w(t) \in \mathcal{W}_N$, the trajectory generated by the interval inputs in Eq. (23c), denoted by x_R , satisfies

$$x_R(t) \in \mathcal{X}_{\text{rec}} \quad \text{for all } t \text{ while recovery remains active.} \quad (24)$$

Together with Eq. (22), this certification preserves physical safety throughout recovery, including any interval after entry into Ω_{ρ^*} during

which nominal handoff remains ineligible. The required availability and finite-time return properties of the certified recovery policy are formalized in the following assumption.

Assumption 7. Suppose that repair remains confirmed and no shutdown request or new abnormal event interrupts recovery before entry into Ω_{ρ^*} . There exists a positive integer N_R such that, for every $x(t_k) \in \mathcal{X}_{\text{rec}}$ and every disturbance signal satisfying $w(t) \in \mathcal{W}_N$ throughout recovery, the certified policy in Eq. (23) generates a trajectory x_R for which there exists $j_R \in \{0, 1, \dots, N_R\}$ satisfying

$$x_R(t_{k+j_R}) \in \Omega_{\rho^*}. \quad (25)$$

This assumption imposes only finite-time return to Ω_{ρ^*} ; it does not prescribe a particular recovery certificate or require V to decrease monotonically along the recovery trajectory. It is therefore a basic recoverability requirement for finite-time entry into the nominal Lyapunov region rather than a stability assumption imposed on the entire recovery trajectory. The assumption is reasonable because it is imposed only after repair has restored the normal disturbance description and only for states in the deliberately selected set $\mathcal{X}_{\text{rec}}$, with both input channels available. It is also necessary: repair alone does not imply that a constrained state can be returned to the region in which the nominal SI-LMPC is certified. Without this recoverability property, the integrated architecture could guarantee containment before repair but could not guarantee entry into the nominal Lyapunov region, which is required for nominal handoff.

At repair confirmation, the pre-repair status update in Eq. (18) and the performance-based intensity rule in Eq. (20) are suspended. The recovery policy in Eq. (23) determines both inputs instead. During recovery, set $q_k = 1$ if $\alpha_k^{\text{rec}} > 0$, and set $q_k = 0$ otherwise, so that q_k records whether the safety-system action is enabled over the current sampling interval. The recovery intensity α_k^{rec} is determined by the recovery law and is distinct from the performance command α_k^{cmd} and the pre-repair applied intensity α_k in Eq. (20). Neither $\alpha_k^{\text{rec}} = 0$ nor $q_k = 0$ clears the supervisory latch. The shutdown conditions specified in Section 3.2 retain priority. If repair confirmation is withdrawn or a new abnormal event is detected, recovery is terminated and the policy in Eq. (21) is resumed whenever its certification conditions hold; otherwise, no nonshutdown guarantee is made unless a separately certified fallback is used. Neither repair confirmation nor these events reset t_{k_a} while the supervisory latch remains active, so the original supervisory deadline continues to apply throughout recovery. A new activation time is assigned only after a completed nominal handoff and a subsequent supervisory activation.

Entry into Ω_{ρ^*} is necessary but not sufficient for return to nominal operation. If such an instant exists, let t_{k_r} be the first sampling instant before $t_{k_a} + t^*$ at which repair remains confirmed, no shutdown condition or new abnormal event is active, $x(t_{k_r}) \in \Omega_{\rho^*}$, $S_{k_r} < S_{\text{th},1}$, and the SI-LMPC problem in Eq. (8) is feasible and attains a finite minimizer. The first nominal interval with $u_s = 0$ must also satisfy the applicable physical safety and nominal eligibility checks. At t_{k_r} , the supervisory latch is reset, $q_{k_r} = 0$ is selected, and $u_{s,k_r} = 0$ is held over $[t_{k_r}, t_{k_r+1})$. Nominal operation then resumes with the SI-LMPC, and its post-handoff safety follows from the nominal result in part 1 of Theorem 1 under the hypotheses stated there.

Because the Safeness Index may depend on measurement history or other safety-relevant information, entry into Ω_{ρ^*} does not imply $S_k < S_{\text{th},1}$. No state-space inclusion between Ω_{ρ^*} and a Safeness Index sublevel set is therefore imposed. If the state reaches Ω_{ρ^*} before the safety information or another handoff condition has cleared, the supervisor remains latched and the recovery policy continues to be applied until nominal handoff or a shutdown request. Consequently, Assumption 7 guarantees finite-time entry into the nominal Lyapunov region under its uninterrupted-recovery hypothesis, but not, by itself, a finite nominal handoff. If the handoff conditions have not been satisfied

before the supervisory deadline, shutdown is requested and nominal operation is not resumed.

Remark 7. Assumption 7 may be verified using reachability analysis, a certified backup controller, a process-specific progress condition, or another suitable method. A Lyapunov decrease condition is one sufficient application-specific construction, but it is not required by the general framework. For example, if $0 < \rho_R < \rho^*$, $\lambda_R > 0$, and a recovery law ensures $\dot{V}(x(t)) + \lambda_R(V(x(t)) - \rho_R) \leq 0$ until sampled entry into Ω_{ρ^*} , then $V(x(t)) - \rho_R \leq \exp(-\lambda_R(t - t_k))(V(x(t_k)) - \rho_R)$, which guarantees finite-time entry and provides an explicit entry-time bound whenever $V(x(t_k)) > \rho^*$. Together with Eq. (24), this verifies the required recovery property. This construction is used in the MIC application.

3.4. Overall safety system overview

The integrated architecture has four operating modes: nominal control, active safety operation, post-repair recovery, and shutdown. At each sampling instant while the supervisory mode is latched, including the activation instant, the supervisor first evaluates the timeout condition $t_k - t_{k_a} \geq t^*$, followed by the shutdown condition $S_k > S_{\text{th},3}$. If neither condition holds, repair confirmation is evaluated. While the abnormal condition remains unresolved, the active-safety policy is applied. Once repair has been confirmed, the recovery policy is applied, and nominal control is resumed only after all handoff conditions have been satisfied. Thus, neither releasing the safety-system input nor entering the nominal Lyapunov region by itself clears the supervisory latch.

During nominal operation, the SI-LMPC in Eq. (8) is applied with $u_s = 0$. The supervisor is activated by the condition in Eq. (9), which accounts for both departure from Ω_{ρ^*} and deterioration of the measured Safeness Index. Once activated, the supervisor remains latched even if the state subsequently re-enters Ω_{ρ^*} or the safety-system input is temporarily returned to zero. Supervisory activation also provides an observable trigger for plant personnel to begin locating and diagnosing the possible abnormal condition and carrying out the required process-side corrective action. Here, repair denotes the physical or operational intervention used to repair or isolate faulty equipment, remove the disturbance source, or reduce its effect sufficiently for $w(t) \in \mathcal{W}_N$ to hold again. It is distinct from the subsequent closed-loop recovery procedure.

Before repair is confirmed, the measurement-based active-safety policy in Eq. (21) is applied without propagating an abnormal process model. While the abnormal condition persists, repeated activation and release of the safety-system input may produce several protective excursions within $\mathcal{X}_A$; convergence to the nominal equilibrium is not required during this phase. For the sampled implementation and the theorem below, select the previously introduced maximum supervisory duration as $t^* = N_f^* \Delta$, where N_f^* is a positive integer. If the supervisory latch remains active at $t_{k_a+N_f^*} = t_{k_a} + t^*$, shutdown is requested regardless of repair status. Repair confirmation neither resets nor suspends this deadline, and cannot initiate recovery or nominal handoff at the deadline.

After repair has been confirmed, the recovery policy in Section 3.3 is applied until the nominal handoff conditions are satisfied or shutdown is requested. During this phase, the supervisory mode remains latched and the shutdown conditions retain priority.

The physical state-space organization is illustrated conceptually in Fig. 1. Only the physical and certified operating sets are represented as state-space regions. The Safeness Index thresholds are sampled supervisory levels and are therefore shown separately rather than as fixed state-space boundaries.

The overall supervisory logic is summarized in Fig. 2. In particular, the figure distinguishes the latched supervisory mode from the

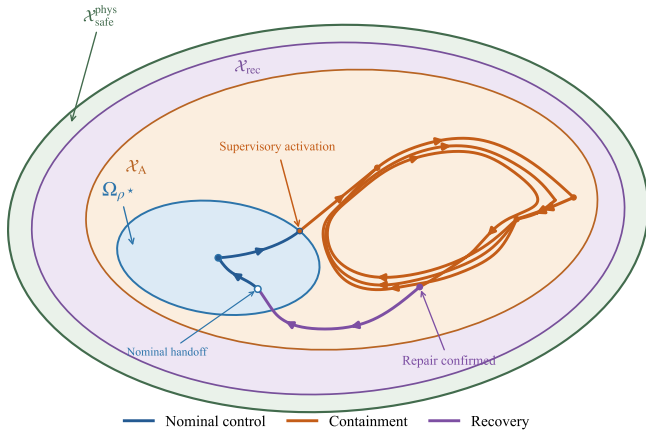


Fig. 1. Illustrative state evolution under the integrated control-safety architecture. Repeated orange excursions represent containment before repair. After repair, the recovery policy returns the state to Ω_{ρ^*} . Nominal operation resumes only after complete handoff.

safety-input activation status q_k used during unresolved active-safety operation. The adjustable intensity in that mode is selected through Eq. (20), rather than as a function of the Safeness Index alone. After repair, the pre-repair status and intensity updates are suspended, and the recovery policy in Eq. (23) first selects $u_{R,k}$ and then determines the supplemental safety-system action. In the figure, q_k denotes the initialized or updated status; the first sampling interval after each safety-input activation during unresolved active-safety operation uses full intensity. The nominal-mode branch requires nominal eligibility and a cleared supervisory latch; return from recovery therefore requires complete handoff before the deadline.

The closed-loop properties supported by the preceding assumptions are summarized next.

Theorem 1. Suppose that Assumptions 1 to 7 hold. Suppose also that the SI-LMPC is feasible and attains a finite minimizer whenever nominal operation is permitted. Then the following properties hold.

1. Suppose nominal operation starts at t_{k_0} , with $x(t_{k_0}) \in \Omega_{\rho^*}$ and $w(t) \in \mathcal{W}_N$ throughout the nominal interval under consideration, and that no supervisory trigger occurs. For the selected sampling period $\Delta > 0$, suppose that there exist $\rho_s > 0$ and $\varepsilon_V > 0$ satisfying

$$L'_x M_F \Delta + L'_w W_{\max} - \beta_3 (\beta_2^{-1}(\rho_s)) \leq -\varepsilon_V. \quad (26)$$

Define

$$\rho_{\min} := \rho_s + M_F \int_0^\Delta \beta_4(\beta_1^{-1}(\rho_s) + M_F \tau) d\tau, \quad (27)$$

and let $\Omega_{\rho_{\min}} := \{x \in D \mid V(x) \leq \rho_{\min}\}$. Suppose further that $0 < \rho_s < \rho_{\min} < \rho^* < \rho$.

Then Ω_{ρ^*} is robustly forward invariant. Define

$$N_N := \left\lceil \frac{\max \left\{ 0, V(x(t_{k_0})) - \rho_{\min} \right\}}{\varepsilon_V \Delta} \right\rceil. \quad (28)$$

There exists $j_N \in \{0, 1, \dots, N_N\}$ such that $x(t_{k_0+j_N}) \in \Omega_{\rho_{\min}}$, and the trajectory remains in $\Omega_{\rho_{\min}}$ thereafter while the nominal mode remains active. Hence, the trajectory remains in $\mathcal{X}_{safe}^{phys}$.

2. For every abnormal event covered by Assumptions 4 to 6, the trajectory remains in $\mathcal{X}_{safe}^{phys}$ from the physical onset of the event to t_{k_a} . It subsequently remains in $\mathcal{X}_A \subset \mathcal{X}_{safe}^{phys}$ throughout every certified nonshutdown active-safety interval. The initialized interval at $k = k_a$ is covered by Assumption 4. For $k \geq k_a + 1$, every interval with

$q_k = 0$ satisfies the zero-input containment condition, which ensures $c(x(t)) \leq 0$. On every such interval with $q_k = 1$ for which at least one $\alpha \in [0, 1]$ satisfies $\hat{r}_k(\alpha) \leq -\eta_c$, $\frac{d}{dt}c(x(t)) \leq -\eta_c \leq 0$ for almost every $t \in [t_k, t_{k+1})$. If no such intensity exists, the applied intensity saturates at $\alpha_k = 1$, and $c(x(t)) \leq 0$ is maintained throughout the interval by Assumptions 5 and 6.

3. During the latched supervisory mode, shutdown is requested at any sampling instant for which $t_k - t_{k_a} \geq t^*$ or $S_k > S_{th,3}$. In particular, if the latch remains active at $t_{k_a+N_f} = t_{k_a} + t^*$, shutdown is requested regardless of repair status and takes priority over recovery or nominal handoff. These statements specify the two shutdown conditions in the proposed supervisory logic and do not certify the shutdown transient.
4. Suppose repair is confirmed at a sampling instant t_{k_f} satisfying $t_{k_f} - t_{k_a} < t^*$, before shutdown during certified active-safety operation. Assume that the confirmation remains valid and that $w(t) \in \mathcal{W}_N$ throughout recovery. Then $x(t_{k_f}) \in \mathcal{X}_A \subseteq \mathcal{X}_{rec}$. If no new abnormal event or shutdown request interrupts recovery before entry and the recovery certificate remains valid, the recovery trajectory x_R , initialized by $x_R(t_{k_f}) = x(t_{k_f})$, remains physically safe while recovery is active, and there exists $j_R \in \{0, 1, \dots, N_R\}$ such that $x_R(t_{k_f+j_R}) \in \Omega_{\rho^*}$. In particular, sampled entry into Ω_{ρ^*} occurs no later than $t_{k_f} + N_R \Delta$. A sufficient timing condition for this bound to precede the supervisory deadline is $t_{k_f} + N_R \Delta < t_{k_a} + t^*$. The entry bound does not override the shutdown conditions, does not imply nominal handoff, and does not require monotonic decrease of V during recovery.
5. If such a sampling instant exists, then at the first instant at which all handoff conditions in Section 3.3 are satisfied, the supervisory latch is reset, the safety-input status is set to $q_k = 0$, and nominal operation resumes with $u_s = 0$. The result in part 1 then applies while nominal eligibility is maintained and no new supervisory trigger occurs. Finite-time entry into Ω_{ρ^*} does not by itself imply a finite handoff, because the Safeness Index may depend on measurement history or other safety-relevant information.

Proof. Consider a nominal sampling interval beginning at t_k , and hold the first SI-LMPC action over $[t_k, t_{k+1})$. Because $x(t_k) \in \Omega_{\rho^*} \subset \text{int}(\Omega_\rho)$, the bounds in Assumption 1 apply at least locally. Let $\bar{\tau} \in (0, \Delta]$ be the first time at which $V(x(t_k + \bar{\tau})) = \rho$, if such a time exists, and set $\bar{\tau} = \Delta$ otherwise. Then $x(t_k + \tau) \in \Omega_\rho$ for every $\tau \in [0, \bar{\tau}]$, and Eq. (6a) gives

$$\|x(t_k + \tau) - x(t_k)\| \leq M_F \tau, \quad \tau \in [0, \bar{\tau}]. \quad (29)$$

The first-action constraint in Eq. (8f), together with Eqs. (5b) and (6c), gives, for almost every such τ ,

$$\dot{V}(x(t_k + \tau)) \leq -\beta_3(\|x(t_k)\|) + L'_x M_F \tau + L'_w W_{\max}, \quad \tau \in [0, \bar{\tau}]. \quad (30)$$

If $V(x(t_k)) \geq \rho_s$, then $\|x(t_k)\| \geq \beta_2^{-1}(\rho_s)$. Hence, Eq. (26) makes the right-hand side of Eq. (30) no greater than $-\varepsilon_V$, and integration yields

$$V(x(t_k + \tau)) \leq V(x(t_k)) - \varepsilon_V \tau, \quad \tau \in [0, \bar{\tau}]. \quad (31)$$

If $V(x(t_k)) < \rho_s$, then $\|x(t_k)\| < \beta_1^{-1}(\rho_s)$. It follows from Eq. (29) that

$$\|x(t_k + s)\| < \beta_1^{-1}(\rho_s) + M_F s, \quad s \in [0, \bar{\tau}]. \quad (32)$$

Using Eqs. (5c) and (6a), one then obtains, for $\tau \in [0, \bar{\tau}]$,

$$\begin{aligned} V(x(t_k + \tau)) - V(x(t_k)) &= \int_0^\tau \frac{\partial V(x(t_k + s))}{\partial x} F(x(t_k + s), u(t_k), 0, w(t_k + s)) ds \\ &\leq M_F \int_0^\tau \beta_4(\beta_1^{-1}(\rho_s) + M_F s) ds. \end{aligned} \quad (33)$$

Therefore, for $\tau \in [0, \bar{\tau}]$,

$$\begin{aligned} V(x(t_k + \tau)) &< \rho_s + M_F \int_0^\Delta \beta_4(\beta_1^{-1}(\rho_s) + M_F s) ds \\ &= \rho_{\min} < \rho^*. \end{aligned} \quad (34)$$

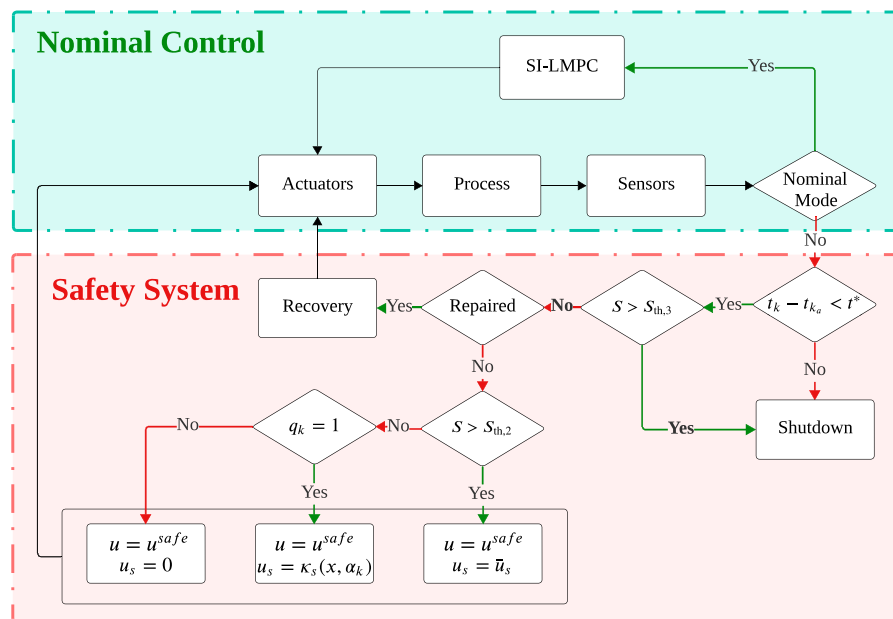


Fig. 2. Sampled logic for nominal control, active safety operation, post-repair recovery, and shutdown.

If $\bar{\tau}$ were the first time at which V reached ρ , then Eq. (31) or Eq. (34) would give $V(x(t_k + \bar{\tau})) < \rho$, contradicting $V(x(t_k + \bar{\tau})) = \rho$. Hence, the trajectory cannot leave Ω_ρ , and all the estimates above hold over the entire sampling interval. They also show directly that a trajectory beginning in Ω_{ρ^*} cannot leave that set. Induction over the sampling instants proves its robust forward invariance. Whenever $V(x(t_k)) > \rho_{\min}$, one has $V(x(t_k)) > \rho_s$, and Eq. (31) gives

$$V(x(t_{k+1})) \leq V(x(t_k)) - \varepsilon_V \Delta. \quad (35)$$

Iteration of Eq. (35) proves entry into $\Omega_{\rho_{\min}}$ within the bound in Eq. (28). Once a sampled state belongs to $\Omega_{\rho_{\min}}$, Eq. (31) applies when $V(x(t_k)) \geq \rho_s$, giving $V(x(t_k + \tau)) \leq V(x(t_k)) \leq \rho_{\min}$, whereas Eq. (34) applies when $V(x(t_k)) < \rho_s$, giving $V(x(t_k + \tau)) < \rho_{\min}$. Thus, the continuous-time trajectory remains in $\Omega_{\rho_{\min}}$ thereafter. This proves part 1.

For part 2, Assumption 4 keeps the trajectory in $\mathcal{X}_{\text{safe}}^{\text{phys}}$ from the physical onset of the abnormal event to t_{k_a} , places $x(t_{k_a})$ in $\mathcal{X}_A$, and certifies containment in $\mathcal{X}_A$ over the initialized first active-safety interval. The first rate estimate used by the policy is obtained at t_{k_a+1} from that completed interval, and Assumption 5 applies for $k \geq k_a + 1$.

Consider a subsequent active-safety interval with $q_k = 0$. For either previous status, the update in Eq. (18) permits $q_k = 0$ only when Eq. (16) holds. Thus, Eq. (17) gives $c(x(t)) \leq 0$.

Now consider a subsequent interval with $q_k = 1$. If at least one $\alpha \in [0, 1]$ satisfies $\hat{r}_k(\alpha) \leq -\eta_c$, the adjustable branch applies $\alpha_k \in [\underline{\alpha}_k, 1]$, whereas the maximum-action branch applies $\alpha_k = 1$. By Eq. (12), both choices satisfy $\hat{r}_k(\alpha_k) \leq -\eta_c$, and Assumption 5 gives $\dot{c}(x(t)) \leq -\eta_c \leq 0$ for almost every time in the interval. If no such intensity exists, the applied intensity saturates at $\alpha_k = 1$, and Assumption 5 and Eq. (19) ensure $c(x(t)) \leq 0$ throughout the interval. Induction therefore keeps the trajectory in $\mathcal{X}_A$ throughout all certified nonshutdown active-safety intervals. This proves part 2.

For part 3, the decision priority in Section 3.2 first evaluates $t_k - t_{k_a} \geq t^*$ while the supervisory latch remains active. If the deadline has not been reached, it evaluates $S_k > S_{\text{th},3}$. Either condition requests shutdown before repair confirmation or nominal handoff can select another mode. Hence, a latched supervisory episode cannot continue beyond $t_{k_a+N_f^*} = t_{k_a} + t^*$, including when repair is confirmed at that instant.

For part 4, part 2 gives $x(t_{k_f}) \in \mathcal{X}_A$ at the repair-confirmation instant. The inclusion in Eq. (22) therefore gives $x(t_{k_f}) \in \mathcal{X}_{\text{rec}}$. While the

recovery certificate remains valid, Eq. (24) keeps the resulting trajectory x_R in $\mathcal{X}_{\text{rec}} \subset \text{int}(\mathcal{X}_{\text{safe}}^{\text{phys}})$. Starting from $x_R(t_{k_f}) = x(t_{k_f})$, Assumption 7 gives some $j_R \in \{0, 1, \dots, N_R\}$ for which $x_R(t_{k_f+j_R}) \in \Omega_{\rho^*}$. Thus, recovery remains physically safe and reaches the target within at most N_R sampling periods, provided that recovery is not interrupted before entry. The corresponding sampled entry bound is $t_{k_f} + N_R \Delta$, and it precedes the supervisory deadline whenever $t_{k_f} + N_R \Delta < t_{k_a} + t^*$. This argument uses only the certified recovery containment and finite-time return properties; it does not require V to decrease monotonically before the target set is reached.

Finally, if and when all handoff conditions in Section 3.3 are satisfied, the supervisor is reset and the SI-LMPC is applied with $u_s = 0$. Part 1 then applies while the nominal eligibility conditions continue to hold and no new supervisory trigger occurs. Because Assumption 7 guarantees entry into Ω_{ρ^*} but does not bound the time required for the history-dependent safety information and the other handoff conditions to clear, this final step does not assert a finite nominal-handoff time. This proves part 5. $\square$

No nonshutdown safety claim is made when the assumptions supporting the active-safety or recovery certificates cease to hold. Likewise, safety of the shutdown transient requires a separate, process-specific analysis.

Remark 8. In practical implementation, the proposed policy should be coordinated with existing alarms, interlocks, and protection systems. Adjustments intended to reduce material loss must not override required protective actions or shutdown requests. Each safety instrumented function, including its sensors, decision logic, and actuators, should be evaluated against applicable functional safety requirements, including the required safety integrity level (SIL).

4. Application to a safety-critical chemical process

4.1. Process description

The application considers the strongly exothermic hydrolysis of MIC in a jacketed continuous stirred-tank reactor (CSTR). MIC is a highly reactive, low-boiling, and highly toxic chemical intermediate and was the principal chemical involved in the Bhopal disaster (Ball, 2011; Toro et al., 2016). Motivated by that accident, previous dynamic

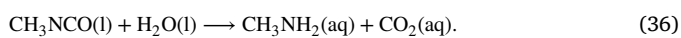
Table 2

Process parameters and nominal operating conditions for the MIC hydrolysis CSTR.

Parameter	Value	Description
M	4.1×10^4 kg	Constant mass of the reacting liquid.
F_r	57.5 kg/s	Inlet and outlet mass flow rate.
T_0	293 K	Inlet-stream temperature.
C_p	3000 J/(kg K)	Specific heat capacity of the reacting liquid.
UA	7.1×10^6 J/(s K)	Overall heat-transfer conductance.
k_0	4.13×10^8 s ⁻¹	Pre-exponential factor.
E_a	6.54×10^4 J/mol	Activation energy.
ΔH	-8.04×10^4 J/mol	Heat of the MIC hydrolysis reaction.
R	8.314 J/(mol K)	Universal gas constant.
C_{A0s}	29.35 mol/kg	Nominal inlet MIC concentration.
C_{As}	10.1767 mol/kg	Nominal reactor MIC concentration.
T_s	305.1881 K	Nominal reactor temperature.
T_{js}	293 K	Nominal cooling-jacket temperature.

and bifurcation studies of MIC hydrolysis have shown that its thermal behavior can be highly sensitive to heat removal and operating conditions. In particular, the process may lose thermal stability through an abrupt transition to large-amplitude temperature oscillations rather than through a gradual approach to a steady-state ignition point (Ball, 2011; Toro et al., 2016). The combination of MIC toxicity and the potential for rapid thermal runaway makes this reaction a representative safety-critical chemical process. The present application does not attempt to reconstruct the storage tank involved in the Bhopal accident. Instead, it adopts the jacketed CSTR model used in Zhang et al. (2018a) to represent the principal concentration–temperature dynamics associated with MIC hydrolysis.

The primary hydrolysis reaction converts MIC and water to methylamine and carbon dioxide according to



MIC may also react with the methylamine produced by Eq. (36) to form dimethyl urea; however, reliable kinetic data for this secondary reaction are unavailable, and previous thermal-instability models retain only the primary hydrolysis reaction (Ball, 2011; Toro et al., 2016). Following these studies, the MIC consumption rate is represented by the first-order Arrhenius expression

$$r_A(C_A, T) = k_0 \exp\left(-\frac{E_a}{RT}\right) C_A, \quad (37)$$

where r_A is the MIC consumption rate in mol/(kg s), C_A is the mass-specific MIC concentration in the reactor, T is the reactor temperature, k_0 is the pre-exponential factor, E_a is the activation energy, and R is the universal gas constant. In this reduced kinetic representation, water does not appear as an independent dynamic variable.

The reactor is assumed to be perfectly mixed and to contain a constant liquid mass M . The inlet and outlet mass flow rates are equal and are denoted by F_r . The inlet stream has MIC concentration C_{A0} and temperature T_0 . Heat is exchanged with a cooling jacket maintained at temperature T_j . The reacting liquid is assumed to have a constant specific heat capacity C_p , and the overall reactor-jacket heat-transfer conductance is denoted by UA . Under these assumptions, the MIC material balance and reactor energy balance are given by

$$M\dot{C}_A = -Mr_A(C_A, T) + F_r(C_{A0} - C_A), \quad (38a)$$

$$MC_p\dot{T} = (-\Delta H)Mr_A(C_A, T) + F_rC_p(T_0 - T) - UA(T - T_j), \quad (38b)$$

where $\Delta H < 0$ is the heat of the hydrolysis reaction. The model parameters and nominal operating conditions are adopted from Zhang et al. (2018a) and are summarized in Table 2.

At $C_{A0} = C_{A0s}$, $T_0 = 293$ K, and $T_j = T_{js}$, the reactor is operated at the steady state $(C_A, T) = (C_{As}, T_s)$. An increase in the inlet MIC concentration raises the MIC inventory and, consequently, the reaction heat-generation rate. Because the reaction rate in Eq. (37) increases

exponentially with temperature, the resulting temperature increase further accelerates MIC consumption and generates additional heat. If this heat-generation rate exceeds the heat removed by the process stream and the cooling jacket, the concentration–temperature coupling forms a positive feedback mechanism that can produce thermal runaway. This strong nonlinearity and sensitivity to heat removal provide the intrinsic safety-critical character of the process.

The model in Eq. (38) is a lumped model with the MIC concentration and reactor temperature as its dynamic variables. If the calculated temperature reaches conditions at which significant vaporization would occur, the trajectory should be interpreted as indicating a severe thermal excursion rather than as a quantitative prediction of the subsequent multiphase behavior.

4.2. Process safety characterization

The principal process hazard considered in this application is thermal runaway. Although the MIC concentration influences the reaction heat-generation rate, the onset of thermal runaway is directly reflected by an elevated reactor temperature and a sustained or accelerating temperature increase (Ball, 2011; Toro et al., 2016; Zhang et al., 2018a). Therefore, the process-specific Safeness Index is constructed using the measured absolute reactor temperature and its recent evolution. This construction does not require knowledge of the disturbance or prediction of a future state trajectory.

Consistent with the control formulation below, define $x_1 = C_A - C_{As}$ and $x_2 = T - T_s$, and collect these deviations as $x = [x_1, x_2]^T$. Let $T_{\text{lim}} > T_s$ be the limiting reactor temperature. The physical thermal safe set is

$$\mathcal{X}_{\text{safe}}^{\text{phys}} = \{x \in D \mid T_s + x_2 \leq T_{\text{lim}}\}. \quad (39)$$

The Safeness Index introduced next is an early-warning measure and is not used as a replacement for this physical constraint.

Let $T_k = T(t_k)$ denote the reactor-temperature measurement at the sampling instant t_k . Select a moving-window length $T_w = N_w\Delta$, where $\Delta > 0$ is the measurement interval and $N_w \in \mathbb{Z}_{>0}$. The temperature increments over the current and preceding windows are defined as

$$\Delta T_k^{\text{cur}} := T_k - T_{k-N_w}, \quad (40a)$$

$$\Delta T_k^{\text{pre}} := T_{k-N_w} - T_{k-2N_w}. \quad (40b)$$

The quantity $2\Delta T_k^{\text{cur}} - \Delta T_k^{\text{pre}}$ is a linear extrapolation of the most recent temperature increment. It becomes greater than ΔT_k^{cur} when the measured temperature increase is accelerating. To ensure that a negative temperature increment does not directly reduce the indicated thermal risk, define

$$\Delta T_k^{\text{risk}} := \max\{0, \Delta T_k^{\text{cur}}, 2\Delta T_k^{\text{cur}} - \Delta T_k^{\text{pre}}\}. \quad (41)$$

For $0 \leq k < N_w$, before a complete current-window history becomes available, both ΔT_k^{cur} and ΔT_k^{pre} are initialized to zero. For $N_w \leq k < 2N_w$, ΔT_k^{cur} is evaluated using Eq. (40a), whereas ΔT_k^{pre} is initialized as ΔT_k^{cur} . For $k \geq 2N_w$, both increments are evaluated using Eqs. (40a) and (40b).

Use T_{lim} to normalize the thermal condition, and define $\text{sat}_{[-1,1]}(z) = \min\{1, \max\{-1, z\}\}$. The process-specific Safeness Index is

$$S_k = \text{sat}_{[-1,1]}\left(\frac{T_k - T_s + \gamma_S \Delta T_k^{\text{risk}}}{T_{\text{lim}} - T_s}\right), \quad (42)$$

where $\gamma_S \geq 0$ is a dimensionless weighting coefficient specifying the contribution of the recent temperature trend. By construction, $S_k \in [-1, 1]$.

To interpret Eq. (42), define the effective thermal condition as $T_k^{\text{eff}} = T_k + \gamma_S \Delta T_k^{\text{risk}}$. The Safeness Index satisfies $S_k = 0$ when $T_k^{\text{eff}} = T_s$. It is negative when $T_k^{\text{eff}} < T_s$ and positive when $T_k^{\text{eff}} > T_s$. Between the lower and upper saturation limits, the Safeness Index varies linearly

with T_k^{eff} . Specifically, $S_k = -1$ when $T_k^{\text{eff}} \leq 2T_s - T_{\text{lim}}$, whereas the limiting value $S_k = 1$ is reached when $T_k^{\text{eff}} \geq T_{\text{lim}}$.

Because $\Delta T_k^{\text{risk}} \geq 0$, a positive or accelerating temperature trend can only increase the Safeness Index. Consequently, a rapid temperature increase may cause S_k to approach its limiting value before the measured temperature itself reaches T_{lim} . For this construction, $S_k < 1$ implies $T_k < T_{\text{lim}}$, but the converse does not necessarily hold because the trend term may produce $S_k = 1$ while $T_k < T_{\text{lim}}$. Thus, Safeness Index levels and the boundary of $\mathcal{X}_{\text{safe}}^{\text{phys}}$ are not interchangeable. If both the current temperature increment and its linear extrapolation are nonpositive, then $\Delta T_k^{\text{risk}} = 0$, and the Safeness Index is determined only by the current absolute temperature. A decreasing current temperature alone does not necessarily imply $\Delta T_k^{\text{risk}} = 0$, because the extrapolated increment may remain positive.

For the simulations considered in this work, temperature measurements are obtained every $\Delta = 1$ s, and the moving-window length is $T_w = 1$ s. Thus, $N_w = 1$, $\Delta T_k^{\text{cur}} = T_k - T_{k-1}$, and $\Delta T_k^{\text{pre}} = T_{k-1} - T_{k-2}$. The remaining parameters are $T_s = 305.1881$ K, $T_{\text{lim}} = 330$ K, and $\gamma_s = 1.10$. The selected value gives the nonnegative temperature-trend contribution a weight slightly above unity; increasing this weight can cause the same Safeness Index threshold to be reached at a lower measured temperature for a given positive trend.

When $\Delta T_k^{\text{risk}} = 0$, a reactor temperature of 320 K corresponds to $S_k = (320 - T_s)/(T_{\text{lim}} - T_s) = 0.59697$. When a positive temperature trend is present, the same Safeness Index level is reached at a lower measured temperature. A smooth temperature-risk measure is used in the nominal SI-LMPC objective, while the measured Safeness Index and its associated threshold levels are used in the subsequent safety-system formulation.

4.3. Nominal control problem

The nominal control objective balances regulation of the MIC concentration and reactor temperature about the steady state (C_{A_s}, T_s) with avoidance of elevated reactor temperature. Using the deviation state defined in Section 4.2, let $u = T_j - T_{js}$ and $w = C_{A0} - C_{A0s}$. Thus, $x = 0$, $u = 0$, and $w = 0$ correspond to the nominal reactor equilibrium.

Let $F_{\text{MIC}}(x, u, w)$ denote the deviation-variable form of the MIC process model in Eq. (38), obtained by setting $C_A = C_{A_s} + x_1$, $T = T_s + x_2$, $T_j = T_{js} + u$, and $C_{A0} = C_{A0s} + w$. The nominal model used for controller prediction is $F_{\text{MIC}}^0(x, u) = F_{\text{MIC}}(x, u, 0)$. Therefore, the prediction model uses the nominal inlet concentration $C_{A0s} = 29.35$ mol/kg, while an unknown change in C_{A0} affects the controller only through subsequent state measurements.

The cooling-jacket temperature is constrained to $280 \text{ K} \leq T_j \leq 300 \text{ K}$. Since $T_{js} = 293 \text{ K}$, the admissible deviation-input set is $U = [-13, 7] \text{ K}$.

Following the Lyapunov-based control design in Zhang et al. (2018a), the quadratic Lyapunov function is selected as

$$V(x) = x^T P x, \quad P = \begin{bmatrix} 200 & 33 \\ 33 & 40 \end{bmatrix}. \quad (43)$$

The nominal Lyapunov region is selected as $\Omega_{\rho^*} = \{x \in \mathbb{R}^2 \mid V(x) \leq \rho^*\}$, where $\rho^* = 7000$. The Lyapunov function and the outer stability region Ω_ρ , with $\rho = 8000$, are adopted from Zhang et al. (2018a). In the present work, the inner operational level is selected as $\rho^* = 7000 < \rho$. The largest reactor temperature in this ellipsoid is $T_s + \sqrt{\rho^*(P^{-1})_{22}} \approx 319.42 \text{ K} < 320 \text{ K}$, which is below $T_{\text{lim}} = 330 \text{ K}$. Hence, $\Omega_{\rho^*} \subset \text{int}(\mathcal{X}_{\text{safe}}^{\text{phys}})$. This geometric inclusion does not impose a bound on the history-dependent Safeness Index. Moreover, the choice $\rho^* = 7000$ is used as the operational boundary in the numerical study; verification of all outer-set and robustness bounds required by Theorem 1 is not claimed for this example.

The nominal dynamics can be written as $F_{\text{MIC}}^0(x, u) = f(x) + g u$, where $f(x) = F_{\text{MIC}}^0(x, 0)$ and $g = [0, U A / (M C_p)]^T$. Define $L_f V(x) =$

$\frac{\partial V(x)}{\partial x} f(x)$ and $L_g V(x) = \frac{\partial V(x)}{\partial x} g$. The Sontag-type control law used to construct the reference stabilizing controller (Lin and Sontag, 1991) is

$$h(x) = \begin{cases} -\frac{L_f V(x) + \sqrt{L_f V(x)^2 + L_g V(x)^4}}{L_g V(x)}, & L_g V(x) \neq 0, \\ 0, & L_g V(x) = 0. \end{cases} \quad (44)$$

Because $h(x)$ may lie outside the admissible input set, the reference controller is defined as $\Phi(x) = \text{proj}_U(h(x))$, where proj_U denotes projection onto U . The reference controller is also used as a fallback action if the online optimization cannot be solved successfully.

The measurement-based Safeness Index in Eq. (42) depends on a finite history of temperature measurements and contains maximum and saturation operations. Directly using this expression along a predicted trajectory would require predicting the future measurement history and would introduce nonsmooth expressions into the optimization problem. The measured index is therefore not propagated through the prediction model. Instead, the nominal SI-LMPC uses the smooth temperature-risk measure

$$\psi_T(T) = \frac{1}{\theta_T} \log \left(1 + \exp \left(\theta_T \frac{T - T_s}{T_{\text{sc}}} \right) \right). \quad (45)$$

Here, $T_{\text{sc}} > 0$ is a temperature scale and $\theta_T > 0$ determines the sharpness of the softplus function. The function ψ_T increases monotonically with the predicted temperature, so penalizing ψ_T^2 discourages predicted motion toward elevated thermal conditions. It is a nominal-model-compatible risk surrogate, not a second definition or a prediction of the measured Safeness Index.

At each sampling instant t_k that satisfies the nominal eligibility conditions in Section 3.1, the nominal SI-LMPC solves

$$\min_{u(\cdot) \in S(\Delta)} \int_{t_k}^{t_k + N\Delta} \left[\tilde{x}(t \mid t_k)^T Q_c \tilde{x}(t \mid t_k) + R_c u(t)^2 + \lambda_T [\psi_T(\tilde{T}(t \mid t_k))]^2 \right] dt \quad (46a)$$

$$\text{s.t. } \dot{\tilde{x}}(t \mid t_k) = F_{\text{MIC}}^0(\tilde{x}(t \mid t_k), u(t)) \quad (46b)$$

$$\tilde{x}(t_k \mid t_k) = x(t_k) \quad (46c)$$

$$u(t) \in U, \quad \forall t \in [t_k, t_k + N\Delta] \quad (46d)$$

$$\begin{aligned} \frac{\partial V(x(t_k))}{\partial x} F_{\text{MIC}}^0(x(t_k), u(t_k)) \\ \leq \frac{\partial V(x(t_k))}{\partial x} F_{\text{MIC}}^0(x(t_k), \Phi(x(t_k))). \end{aligned} \quad (46e)$$

In Eq. (46a), $\tilde{T}(t \mid t_k) = T_s + \tilde{x}_2(t \mid t_k)$ is the predicted absolute reactor temperature. The numerical parameters are selected as $Q_c = \text{diag}\{3, 5\}$, $R_c = 100$, $\lambda_T = 1000$, $T_{\text{sc}} = 2 \text{ K}$, and $\theta_T = 2$.

The sampling period is $\Delta = 1$ s, and the prediction horizon contains $N = 10$ sampling periods, resulting in a total prediction time of 10 s. No additional state path constraint is imposed in this implementation. Instead, Eq. (46e) requires the nominal Lyapunov derivative generated by the first SI-LMPC action to be no greater than that generated by the reference controller Φ . Because $\Phi(x(t_k)) \in U$, selecting the reference action for the first sampling period satisfies this Lyapunov constraint with equality. Since no state path constraint is imposed, completing the remaining horizon with any admissible input sequence gives a feasible candidate for the full optimization problem.

Let $u^*(\cdot \mid t_k)$ denote a finite optimal solution of Eq. (46). The receding-horizon implementation applies $u(t) = u^*(t_k \mid t_k)$ over $[t_k, t_{k+1})$ and resolves the optimization problem using the new state measurement at t_{k+1} . In the numerical implementation, the prediction model is discretized using the explicit Euler method with an integration step of 0.05 s, and the resulting nonlinear program is solved using IPOPT through CasADi. If the optimization does not return a finite admissible solution, the reference action $\Phi(x(t_k))$ is applied.

The computational time of the complete SI-LMPC calculation was measured during nominal operation. The average and maximum measured computation times were 67.17 ms and 145.58 ms, respectively,

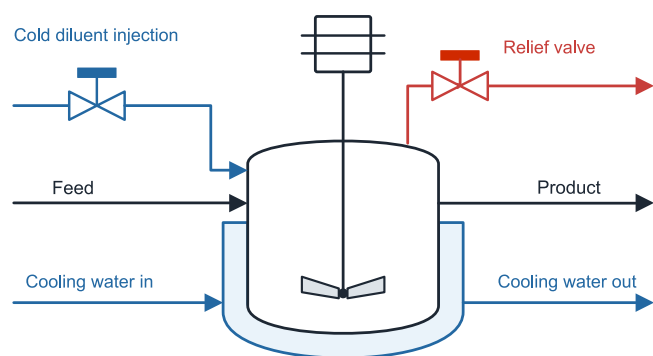


Fig. 3. Schematic of the jacketed MIC hydrolysis CSTR equipped with cold-liquid injection and emergency relief.

both below the sampling period of $\Delta = 1$ s. During active safety operation, the implementation evaluates the measurement-based safety conditions and protective intensity through direct calculations without solving an MPC optimization problem. These measurements characterize controller computation on the tested hardware. Practical implementation must also account for sensing, communication, and actuator response delays, which are not included in this benchmark. If the total implementation time requires a longer sampling period, Δ should be adjusted and the corresponding safety and stability conditions rechecked.

The nominal SI-LMPC is used only while the nominal eligibility conditions remain satisfied. Once the supervisory mode is activated, the protective operating logic introduced in the following subsection takes precedence.

4.4. Safety system

The jacketed MIC hydrolysis reactor is equipped with two physical safety actuators: an emergency relief valve and a cold-liquid injection system. Accordingly, the safety-system input is represented as $u_s = [\dot{m}_{\text{rel}}, \dot{m}_{\text{inj}}]^T$, where $\dot{m}_{\text{rel}}$ and $\dot{m}_{\text{inj}}$ denote the relief-discharge and cold-liquid injection mass flow rates, respectively. As illustrated in Fig. 3, the relief valve removes material from the well-mixed reactor, whereas the injected stream provides dilution and sensible cooling.

The temperature-dependent relief capacity is represented by a surrogate based on a low-vapor-quality homogeneous equilibrium model (HEM) following the emergency vent correlations in Leung (1986) and Hace (2013). Because pressure and vapor quality are not dynamic states in the reduced reactor model, the MIC vapor pressure is used only to construct a temperature-dependent effective discharge capacity. The vapor pressure is represented by $P_v(T) = \exp(a_v - b_v/T)$, with the numerical value of P_v evaluated in Pa, T in K, $a_v = 23.4863168$, and $b_v = 3735.1251$ K. These coefficients reproduce the reported MIC vapor pressure of 348 mmHg at 20°C and predict approximately 755.8 mmHg at the reported boiling point of 39°C (Dhara and Dhara, 2002). The use of an Antoine-type relation is also consistent with the vapor-liquid equilibrium study in Emelianova and Gor (2021); however, the coefficients used here are calibrated for the present reduced model and are not taken directly from that reference. Differentiation gives $dP_v/dT = P_v(T)b_v/T^2$.

Using SI units, the full-capacity flow is calculated with the effective discharge area $A_{\text{eff}} = 10$ m² and the dimensionless coefficient $K_{\text{HEM}} = 0.9003682532$. The two actuators are synchronized through the normalized safety-action intensity $\alpha \in [0, 1]$, resulting in the process-specific map

$$u_s = \kappa_s(x, \alpha) = \alpha \dot{m}_{\text{max}}(T) \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad (47)$$

$$\dot{m}_{\text{max}}(T) = A_{\text{eff}} K_{\text{HEM}} \frac{dP_v}{dT} \sqrt{\frac{T}{C_p}}.$$

Here, $\dot{m}_{\text{max}}$ is calculated using Hace (2013), and $\dot{m}_{\text{rel}} = \dot{m}_{\text{inj}} = \alpha \dot{m}_{\text{max}}(T)$. The two physical flows are not manipulated independently in this application. Both actuators are inactive when $\alpha = 0$, whereas $\alpha = 1$ represents the modeled maximum paired relief and injection action. The effective area A_{eff} is a lumped capacity parameter used in the numerical study and should not be interpreted as a mechanical valve-sizing recommendation. At a fixed reactor temperature, this parameter scales the available maximum relief and injection mass flow rates, thereby affecting the protective capacity and the intensity needed to satisfy the safety condition.

In the numerical implementation, the normalized command is held as $\alpha(t) = \alpha_k$ over $t \in [t_k, t_{k+1})$, whereas the physical flow is evaluated as $u_s(t) = \kappa_s(x(t), \alpha_k)$. Consequently, the relief and injection flow rates may vary within the sampling interval as the reactor temperature changes. This variation must be accounted for when the margin ε_c in Assumption 5 is selected and validated.

Injection of a cold quenchant or diluent is an established approach for mitigating thermal runaway, although its effectiveness depends strongly on the injection and mixing conditions (Hristov and Mann, 2002; Torré et al., 2008). In the present reduced-order model, the injected stream is represented as a cold, MIC-free liquid entering at $T_{\text{inj}} = 280$ K. Anhydrous *o*-dichlorobenzene (*o*-DCB) is adopted solely as an illustrative candidate for this nonaqueous stream. Liquid-phase PVT measurements for 1,2-dichlorobenzene include temperatures near the selected injection temperature (Eastale et al., 1997). This property information does not establish MIC-*o*-DCB chemical compatibility or validate *o*-DCB as an emergency injection medium for MIC. Practical implementation would require separate reactive compatibility, calorimetric, phase-behavior, mixing, and materials-compatibility evaluations.

To remain consistent with the constant-mass assumption in Eq. (38), the injection mass flow is set equal to the relief-discharge mass flow. The discharged material is assumed to have the instantaneous well-mixed reactor composition and temperature. The injected stream contains no MIC, $C_{A,\text{inj}} = 0$, and is assigned the same effective heat capacity as the reactor contents, $C_{P,\text{inj}} = C_P$. Relative to the safety-inactive process model, the contribution of the paired safety action is

$$\left[\begin{array}{c} \dot{C}_A \\ \dot{T} \end{array} \right] \Big|_{\text{safety}} = \frac{\alpha \dot{m}_{\text{max}}(T)}{M} \left[\begin{array}{c} -C_A \\ T_{\text{inj}} - T \end{array} \right]. \quad (48)$$

The first component represents the reduction in reactor MIC concentration caused by replacing discharged reactor material with an MIC-free stream. The second component represents the sensible cooling obtained by replacing hot reactor material with the cold injected liquid.

The reduced-order safety model does not explicitly describe pressure dynamics, vapor quality, latent-heat effects, downstream back pressure, liquid-liquid phase separation, or MIC partitioning between aqueous and organic phases. Consequently, Eq. (47) supplies a temperature-dependent effective flow capacity, while Eq. (48) specifies how the paired action affects the concentration and temperature dynamics.

To apply the active-safety construction in Section 3.2, let $T_A = 329.5$ K $< T_{\text{lim}}$ and $C_{A,\text{max}}^{\text{cert}} = 100$ mol/kg. The compact certification domain, the physical safety certificate, and the known incremental effect of the safety input are defined as

$$D_A := \left\{ x \in \mathbb{R}^2 \mid 0 \leq C_A + x_1 \leq C_{A,\text{max}}^{\text{cert}}, \quad T_{\text{inj}} \leq T_s + x_2 \leq T_{\text{lim}} \right\} \subset D,$$

$$c(x) := T_s + x_2 - T_A, \quad \mathcal{X}_A := \{ x \in D_A \mid c(x) \leq 0 \},$$

$$g_s(x, \kappa_s(x, \alpha)) = \frac{\alpha \dot{m}_{\text{max}}(T)}{M} (T_{\text{inj}} - T) \leq 0, \quad T = T_s + x_2. \quad (49)$$

The selected sets satisfy $\Omega_{\rho^*} \subset \text{int}(\mathcal{X}_A)$ and $\mathcal{X}_A \subset \text{int}(\mathcal{X}_{\text{safety}}^{\text{phys}})$. Moreover, the expression for g_s is determined by the physical actuator model, is independent of the unknown inlet-concentration disturbance, and becomes no larger as α increases. It therefore has the properties required by Assumption 3 on D_A .

The three Safeness Index thresholds are parameterized using the trend-free reference temperatures 320 K, 326 K, and 329 K. Substitution into Eq. (42) with $\Delta T_k^{\text{risk}} = 0$ gives $S_{\text{th},1} = 0.59697$, $S_{\text{th},2} = 0.83879$, and $S_{\text{th},3} = 0.95970$, respectively. These temperatures only parameterize the dimensionless thresholds. All online decisions use the measured Safeness Index, so a positive or accelerating temperature trend can cause a threshold to be reached at a lower measured temperature. The limiting physical temperature remains $T_{\text{lim}} = 330$ K, at which the Safeness Index reaches its limiting value when the trend contribution is zero.

Once Eq. (9) is satisfied, the supervisory mode remains latched and the nominal SI-LMPC is no longer used. The protective feedback law is selected as the maximum available jacket-cooling action, $u^{\text{safe}}(x) = T_{j,\text{min}} - T_{js} = -13$ K, corresponding to $T_j = 280$ K. This input remains applied even when the paired relief and injection action is temporarily inactive. The supervisory latch determines whether nominal control is permitted, whereas q_k indicates whether the paired safety-action branch is enabled during the current sampling interval.

At supervisory activation t_{k_a} , provided that repair has not been confirmed and no shutdown guard is active, the first protective interval uses u^{safe} . If $S_{k_a} \geq S_{\text{th},1}$, the paired action is initialized with $q_{k_a} = 1$ and $\alpha_{k_a} = 1$; otherwise, activation is due to $V(x(t_{k_a})) > \rho^*$, and $q_{k_a} = 0$, $\alpha_{k_a} = 0$. Containment during this first interval is covered by Assumption 4. The first active-safety estimate is calculated at t_{k_a+1} from the completed protective interval. Thereafter, $\hat{d}_k^0$ and $\hat{r}_k(\alpha)$ are updated using Eqs. (13) and (15). For the certificate in Eq. (49), these calculations use successive temperature measurements and the known preceding safety action, without using the unknown inlet concentration or propagating an abnormal-process model.

The MIC implementation uses $\varepsilon_c = 0.22 \text{ K s}^{-1}$, $\eta_c = 0$, and the saturated required intensity $\underline{\alpha}_k$ defined in the general active-safety policy. Once the rate estimate is available, if the feasible-intensity set is nonempty and $q_k = 1$, the selected intensity satisfies $\hat{r}_k(\alpha_k) \leq 0$. Under Assumption 5, the condition $\hat{r}_k(\alpha_k) \leq 0$ ensures that the reactor temperature is nonincreasing over the sampling interval because $c(x) = T - T_A$. The numerical implementation requests shutdown when the required active-safety rate condition is infeasible, rather than continuing through the maximum-action containment branch of the general framework. It also requests shutdown if no admissible intensity satisfies the interval containment condition. During unresolved active-safety operation, whenever $q_k = 0$ after initialization, safety under the protective jacket-cooling input is certified by Eq. (16).

After the first protective interval, during unresolved active-safety operation without an active shutdown guard, the paired safety action is enabled from $q_{k-1} = 0$ if $S_k \geq S_{\text{th},1}$ or Eq. (16) fails, as specified in Eq. (18). When $q_{k-1} = 1$, provided that repair has not been confirmed and no shutdown guard is active, the release rule is

$$q_k = \begin{cases} 0, & S_k < 0, \quad \hat{r}_k(0) \leq 0, \\ & \text{and Eq. (16) holds,} \quad \text{when } q_{k-1} = 1. \\ 1, & \text{otherwise,} \end{cases} \quad (50)$$

This hysteresis keeps the safety-action branch enabled until all release conditions are satisfied. Releasing the safety input does not clear the supervisory latch, and the protective jacket-cooling action remains active.

The MIC implementation applies $\alpha_k = 1$ for one sampling interval following every transition from $q_{k-1} = 0$ to $q_k = 1$, including initialization with $q_{k_a} = 1$. For a sustained episode, maximum action is retained when $S_k > S_{\text{th},2}$. When $q_k = 1$ and $S_k \leq S_{\text{th},2}$, the sustained adjustable branch uses Eq. (20), giving $\alpha_k = \max\{\underline{\alpha}_k, \alpha_j^{\text{cmd}}\}$. The infeasibility shutdown checks take priority over this intensity selection. Here, $\alpha_j^{\text{cmd}} \in [0, 1]$ is the cycle-to-cycle command associated with the current evaluation cycle.

For comparison, the binary safety system uses the same supervisory activation, protective feedback law, status q_k , release condition, and

shutdown guards. Its intensity is restricted to $\alpha_k = 1$ whenever $q_k = 1$ and $\alpha_k = 0$ whenever $q_k = 0$. Thus, the binary and smart cases differ only during the sustained adjustable-action branch.

To evaluate the cycle-to-cycle performance, let k_j denote the sampling index of the j th transition satisfying $q_{k_j-1} = 0$ and $q_{k_j} = 1$, and let $\sigma_j = t_{k_j}$ be the corresponding safety-action activation time. The complete evaluation interval, its duration, and the discharged material are defined by

$$\mathcal{E}_j := [\sigma_j, \sigma_{j+1}), \quad \tau_j^{\text{cyc}} := \sigma_{j+1} - \sigma_j, \quad J_j := \int_{\sigma_j}^{\sigma_{j+1}} \dot{m}_{\text{rel}}(t) dt. \quad (51)$$

Here, J_j is the total material discharged during the j th complete evaluation interval, whereas J_j/τ_j^{cyc} is the corresponding cycle-normalized material-loss metric used for performance evaluation and comparison. Intervals during which the relief valve is closed contribute zero to J_j but remain part of τ_j^{cyc} . Thus, each complete evaluation interval contains one safety-action episode and the subsequent safety-inactive period, if any, before the next activation. The term cycle is used only to define a repeatable evaluation interval and does not imply that the closed-loop trajectory is periodic.

The command α_j^{cmd} is adjusted through a projected pattern search using only completed measurements of the cycle-normalized material-loss metric J_j/τ_j^{cyc} . The search starts from $\alpha_j^{\text{cmd}} = 1$ with an initial step of 0.25. After each command change, one complete evaluation cycle is excluded as a transition cycle, and the median of the next three comparable values of J_j/τ_j^{cyc} is used to evaluate the candidate. Let R^{best} denote the incumbent median value of this metric. A candidate is accepted if its median score is smaller than R^{best} , subject to numerical tolerance. Otherwise, the search direction is reversed and the step is halved. Candidate commands are projected onto $[0, 1]$. The search stops when a zero command is accepted or when the step falls below 0.005, retaining the best accepted command.

Candidate scores use the actual discharged material, including the effects of the saturated required intensity, the prescribed one-sample maximum action following an ordinary reactivation, and maximum action triggered by the Safeness Index threshold. These actions are part of the candidate policy and do not by themselves invalidate its evaluation. An evaluation interrupted by repair or shutdown, or affected by a departure from the certification domain or reactivation within 5 s after release, is not used to compare commands. Excluded evaluation cycles remain part of the physical material-loss accounting. The resulting command is interpreted as the best command accepted by the projected search; no claim of global optimality is made.

The MIC implementation follows the supervisory logic in Sections 3.2 and 3.4, with the additional infeasibility shutdown checks described above. Nominal control is resumed only after repair has been confirmed and the handoff conditions in Section 3.3 are satisfied. While the supervisor remains latched, shutdown is requested if $t_k - t_{k_a} \geq t^*$ or $S_k > S_{\text{th},3}$. The time limit covers the entire supervisory period, including recovery; repair confirmation does not reset or suspend the timer.

Once repair has been confirmed, the unresolved-disturbance activation and release policy and the cycle-to-cycle adjustment are suspended, but the safety input remains available under the recovery law and the supervisory mode remains latched. For the recovery scenario, repair is scheduled at $t_{k_f} = 3600$ s, at which time the inlet MIC concentration is returned from 70 mol/kg to its nominal value $C_{A0s} = 29.35$ mol/kg. The process state is not reset. Therefore, the abnormal disturbance is removed while the state remains at the condition reached immediately before repair. The repaired dynamics used during recovery are

$$\dot{x} = F_{\text{MIC}}^0(x, u) + \frac{\alpha \dot{m}_{\text{max}}(T_s + x_2)}{M} \begin{bmatrix} -(C_{As} + x_1) \\ T_{\text{inj}} - (T_s + x_2) \end{bmatrix}. \quad (52)$$

The recovery scenario assumes that the inlet disturbance is removed completely after t_{k_f} . If a residual disturbance $w(t) \in \mathcal{W}_N$ is retained in

another application, its effect must be included in the recovery margins described below.

The recovery design first computes a continuous feedback control action and then uses the safety-system input only to supply the remaining recovery capability. Using the Lie derivatives $L_f V$ and $L_g V$ defined in Section 4.3, let $a_k = L_f V(x(t_k))$ and $b_k = L_g V(x(t_k))$. With the numerical tolerance $\varepsilon_g = 10^{-10}$, the candidate feedback control action selected at $t_k \geq t_{k_f}$ is

$$u_{R,k} = \begin{cases} \text{proj}_U \left(-\frac{b_k^3}{\sqrt{a_k^2 + b_k^4} - a_k} \right), & a_k < 0, \quad |b_k| > \varepsilon_g, \\ \text{proj}_U \left(-\frac{a_k + \sqrt{a_k^2 + b_k^4}}{b_k} \right), & a_k \geq 0, \quad |b_k| > \varepsilon_g, \\ 0, & |b_k| \leq \varepsilon_g. \end{cases} \quad (53)$$

The first two branches are algebraically equivalent to the standard Sontag formula (Lin and Sontag, 1991). The first form avoids numerical cancellation when $L_f V(x(t_k)) < 0$ and $L_g V(x(t_k))$ approaches zero. This Sontag-based implementation replaces a bound-selection rule based on the sign of $L_g V$, which would cause u to switch between its lower and upper limits as $L_g V$ changes sign. The projection is required because the jacket temperature deviation must satisfy $u_{R,k} \in U = [-13, 7]$ K. All subsequent recovery calculations use the projected value of $u_{R,k}$ as the candidate feedback control action.

Select an interior Lyapunov level $\rho_R = 0.90\rho^* = 6300$, a recovery rate $\lambda_R = 0.008 \text{ s}^{-1}$, and the sampling margin $m_\Delta = 60$ in the numerical units of $\dot{V}$. The level ρ_R is an interior anchor for the recovery condition and is not the nominal handoff boundary. The choice $\rho_R < \rho^*$ requires a strictly negative Lyapunov derivative at $V = \rho^*$, supporting finite-time entry provided that the recovery condition remains satisfied. At each recovery sampling instant, the inputs are required to satisfy

$$\frac{\partial V(x(t_k))}{\partial x} \left[F_{\text{MIC}}^0(x(t_k), u_{R,k}) + \frac{\alpha_k \dot{m}_{\max}(T_s + x_2(t_k))}{M} \begin{bmatrix} -(C_{As} + x_1(t_k)) \\ T_{\text{inj}} - (T_s + x_2(t_k)) \end{bmatrix} \right] + \lambda_R (V(x(t_k)) - \rho_R) + m_\Delta \leq 0. \quad (54)$$

The margin m_Δ reserves recovery authority for changes in the Lyapunov residual over an interval with $u(t) = u_{R,k}$ and $\alpha(t) = \alpha_k$ held constant, while $u_s(t) = \kappa_s(x(t), \alpha_k)$ varies with the state. Specifically, the required bound is that the increase of $\dot{V}(x(t)) + \lambda_R(V(x(t)) - \rho_R)$ from its value at t_k along this trajectory is no greater than m_Δ for all $t \in [t_k, t_{k+1})$. When this bound is valid, Eq. (54) ensures that $\dot{V}(x(t)) + \lambda_R(V(x(t)) - \rho_R) \leq 0$ throughout the sampling interval.

Because Eq. (54) is affine in α_k , direct rearrangement gives the following lower bound when the denominator below is strictly positive, which means that the full safety action is a strict Lyapunov descent direction at $x(t_k)$:

$$\underline{\alpha}_{V,k} := \frac{\frac{\partial V(x(t_k))}{\partial x} F_{\text{MIC}}^0(x(t_k), u_{R,k}) + \lambda_R (V(x(t_k)) - \rho_R) + m_\Delta}{-\frac{\partial V(x(t_k))}{\partial x} \frac{\dot{m}_{\max}(T_s + x_2(t_k))}{M} \begin{bmatrix} -(C_{As} + x_1(t_k)) \\ T_{\text{inj}} - (T_s + x_2(t_k)) \end{bmatrix}}, \quad (55)$$

$$\alpha_k \geq \underline{\alpha}_{V,k}.$$

If $\underline{\alpha}_{V,k} \leq 0$, the selected feedback control action already satisfies the Lyapunov condition, so this condition imposes no positive lower bound on α_k .

The physical certificate is retained during recovery. In addition to Eq. (54), the implementation imposes the sampled temperature guard

$$c(x(t_k)) + \Delta \left[F_{\text{MIC}}^0(x(t_k), u_{R,k}) \right]_2 + \frac{\alpha_k \dot{m}_{\max}(T_s + x_2(t_k))}{M} (T_{\text{inj}} - T_s - x_2(t_k)) \leq 0, \quad (56)$$

where $[\cdot]_2$ denotes the second component of a vector. This is an additional frozen-derivative guard and does not replace the continuous-time checks of D_A , $\mathcal{X}_A$, and $T < T_{\text{lim}}$.

Using $c(x(t_k)) = T_s + x_2(t_k) - T_A$, dividing Eq. (56) by $\Delta > 0$, and noting that $T_s + x_2(t_k) > T_{\text{inj}}$ and $\dot{m}_{\max}(T_s + x_2(t_k)) > 0$ during recovery, the temperature condition gives

$$\underline{\alpha}_{T,k} := \frac{\frac{T_s + x_2(t_k) - T_A}{\Delta} + [F_{\text{MIC}}^0(x(t_k), u_{R,k})]_2}{\frac{\dot{m}_{\max}(T_s + x_2(t_k))}{M} (T_s + x_2(t_k) - T_{\text{inj}})}, \quad (57)$$

$$\alpha_k \geq \underline{\alpha}_{T,k}.$$

If $\underline{\alpha}_{T,k} \leq 0$, the temperature condition imposes no positive lower bound on the safety-action intensity.

After $u_{R,k}$ has been fixed and when the two denominators in Eqs. (55) and (57) are strictly positive, combining the two lower bounds with $\alpha_k \geq 0$ gives

$$\alpha_k = \max \left\{ 0, \underline{\alpha}_{V,k}, \underline{\alpha}_{T,k} \right\}. \quad (58)$$

The bound $\underline{\alpha}_{V,k}$ supplies the additional action needed to satisfy the Lyapunov recovery condition, whereas $\underline{\alpha}_{T,k}$ supplies the action needed by the temperature guard. A negative lower bound does not impose a positive safety-action requirement because zero is included in the maximum.

The positive denominators defining $\underline{\alpha}_{V,k}$ and $\underline{\alpha}_{T,k}$ correspond to negative safety-action directions in the original inequalities. If either denominator is nonpositive, the corresponding ratio is not evaluated and feasibility is determined directly from Eqs. (54) and (56). A positive residual with a nonpositive rearranged coefficient is infeasible. A nonpositive residual with a positive original safety-action direction instead places an upper bound on α_k , which must be respected if the other condition requires positive intensity. The computed intensity is therefore accepted only if it is finite, satisfies $0 \leq \alpha_k \leq 1$, and satisfies Eqs. (54) and (56) after direct substitution. If the required value exceeds one, it is not silently truncated to the available maximum; the recovery calculation is declared infeasible. In this case, no nonshutdown recovery guarantee is made unless a separately certified fallback is used.

Because the feedback control action is selected before α_k , under the stated positive-coefficient conditions Eq. (58) gives the smallest safety intensity satisfying the two recovery conditions for the fixed $u_{R,k}$. If $\alpha_k = 0$, the selected feedback control action is sufficient and recovery proceeds with $u_s = 0$. If $\alpha_k > 0$, the safety-system input is $u_s(t) = \kappa_s(x(t), \alpha_k)$ and supplements $u_{R,k}$ over the same sampling interval. This is not a claim that α_k is the global minimum over every joint selection of $u \in U$ and $\alpha \in [0, 1]$. No grid search, model rollout, or online optimization is used to allocate the recovery inputs. Both $u_{R,k}$ and α_k are recalculated at every recovery sampling instant. Thus, $\alpha_k = 0$ at one instant does not permanently deactivate the safety input; it can be reintroduced later if the feedback control action alone no longer satisfies the recovery conditions.

Provided that the recovery calculation remains feasible and the sampling margin is valid, the implemented recovery inputs satisfy

$$\dot{V}(x(t)) + \lambda_R (V(x(t)) - \rho_R) \leq 0 \quad (59)$$

throughout recovery. Since $\lambda_R > 0$, multiplication by the positive integrating factor $\exp(\lambda_R(t - t_{k_f}))$ rewrites this inequality as the first

relation below. Integrating it from t_{k_f} to t gives the second relation:

$$\frac{d}{dt} \left[\exp(\lambda_R(t - t_{k_f})) (V(x(t)) - \rho_R) \right] \leq 0, \quad (60)$$

$$V(x(t)) - \rho_R \leq \exp(-\lambda_R(t - t_{k_f})) \times (V(x(t_{k_f})) - \rho_R).$$

For $V(x(t_{k_f})) > \rho^* > \rho_R$, provided that the certified recovery procedure remains active until entry, Eq. (60) guarantees entry into Ω_{ρ^*} . Requiring the right-hand side of the exponential bound to be no greater than $\rho^* - \rho_R$ and solving for the elapsed time show that the first continuous-time entry occurs no later than

$$\frac{1}{\lambda_R} \log \left(\frac{V(x(t_{k_f})) - \rho_R}{\rho^* - \rho_R} \right) \quad (61)$$

after repair confirmation. Because the inputs are updated every Δ , the corresponding sampled elapsed-time bound is

$$\Delta \left\lceil \frac{1}{\lambda_R \Delta} \log \left(\frac{V(x(t_{k_f})) - \rho_R}{\rho^* - \rho_R} \right) \right\rceil. \quad (62)$$

If $V(x(t_{k_f})) \leq \rho^*$, the entry duration is zero and the positive-time expression in Eq. (61) is not used. These bounds concern entry into Ω_{ρ^*} , rather than completion of nominal handoff. For the sampled bound to certify entry before the supervisory deadline, its value must be smaller than the remaining time $t_{k_a} + t^* - t_{k_f}$, and recovery must remain uninterrupted with valid certificates.

The recovery policy supersedes the unresolved-disturbance status rule for q_k , but it does not clear the supervisory latch. During recovery, $q_k = 1$ when $\alpha_k > 0$ and $q_k = 0$ when $\alpha_k = 0$. Shutdown conditions retain the highest priority. If repair confirmation is withdrawn or a new abnormal condition is detected, recovery is terminated and the active-safety policy is resumed whenever its measurement-based certification conditions remain valid. Otherwise, no nonshutdown guarantee is made unless a separately certified fallback is used.

Entry into Ω_{ρ^*} is necessary but not sufficient for return to nominal operation. At a sampling instant satisfying $V(x(t_k)) \leq \rho^*$ and $S_k < S_{th,1}$, the implementation first requires the SI-LMPC problem in Eq. (46) to return a finite admissible solution. The first SI-LMPC action is then checked with $\alpha = 0$ over one sampling interval using the repaired model. Handoff is accepted only if the predicted trajectory remains in $\mathcal{X}_A$, the endpoint remains in Ω_{ρ^*} , and the Safeness Index recomputed from the stored temperature history and the predicted endpoint remains below $S_{th,1}$. These model-based checks are used only after repair, when the nominal model is again considered valid.

When all handoff conditions are satisfied before the supervisory deadline and no shutdown guard is active, the supervisory latch is reset, $q_k = 0$, $\alpha_k = 0$, and the first SI-LMPC input is applied. If the state has entered Ω_{ρ^*} but any handoff condition fails, the supervisor remains latched and the certified recovery law continues subject to the shutdown guards. Reaching the deadline while still latched requests shutdown even if repair has been confirmed or $x(t_k) \in \Omega_{\rho^*}$. After hand-off, the nominal eligibility conditions are evaluated at every sampling instant; the supervisor is reactivated whenever Eq. (9) is satisfied.

4.5. Simulation results

All simulations are initialized at the nominal equilibrium, $x(0) = 0$. At $t = 0$, the inlet MIC concentration of the simulated plant is changed from $C_{A0s} = 29.35$ mol/kg to $C_{A0} = 70$ mol/kg, corresponding to $w = 40.65$ mol/kg. In the containment simulations reported first, the inlet concentration is held at this value throughout the simulated horizon. The nominal prediction model continues to use C_{A0s} , and the safety logic is not supplied with the disturbed feed concentration. The

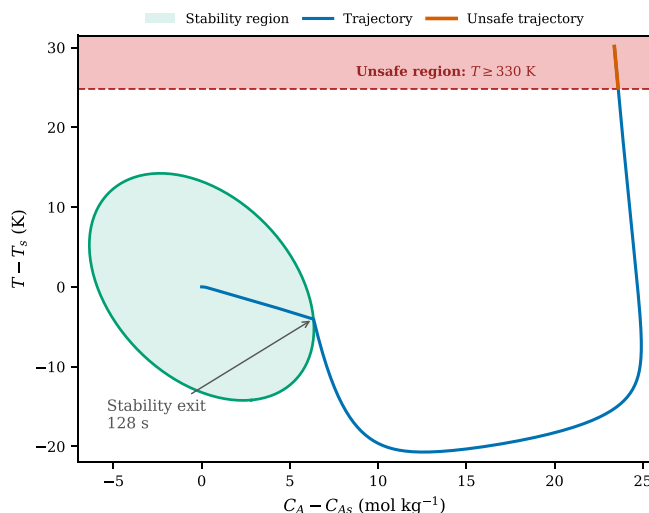


Fig. 4. Thermal runaway illustration of the state space response to the persistent abnormal disturbance with the paired relief and injection action disabled, $u_s = 0$, and execution of the proposed shutdown request disabled. The trajectory leaves Ω_{ρ^*} at $t = 127.9$ s, the supervisor is activated at $t = 128$ s, and the trajectory eventually crosses the physical temperature limit $T_{lim} = 330$ K.

controller sampling period is $\Delta = 1$ s, and the plant is integrated with a 0.01 s step. The maximum supervisory duration is set to $t^* = 15$ h, measured from supervisory activation and including recovery after repair, and lies beyond the corresponding simulated horizons. A separate 1.5 h simulation with repair is reported at the end of this subsection to evaluate recovery and nominal handoff.

To demonstrate the need for the additional safety input, the disturbance is first simulated with the paired relief and injection action disabled, $u_s = 0$. The nominal SI-LMPC is used while the nominal eligibility conditions remain satisfied. Following supervisory activation, the protective feedback action is applied and drives the jacket temperature to its lower bound of 280 K. Thus, feedback cooling remains available in this case even though the physical safety input is disabled. For this diagnostic comparison, execution of the proposed shutdown request is disabled so that the consequence of relying on feedback cooling alone can be observed.

As shown in Fig. 4, the state leaves Ω_{ρ^*} at approximately $t = 127.9$ s, and the supervisor is activated at the next sampling instant, $t = 128$ s. The reactor temperature subsequently decreases to approximately 284.5 K at $t = 235.0$ s, while the MIC concentration continues to increase. The accumulated reactive material eventually causes the trajectory to turn toward higher temperatures. Despite the maximum available jacket cooling, the reactor temperature crosses $T_{lim} = 330$ K at approximately $t = 588.7$ s, as shown in Fig. 5. Therefore, the temporary temperature decrease does not represent recovery, and the feedback control input alone is insufficient to contain this tested disturbance.

The binary and smart safety systems are subsequently evaluated under the same initial condition and persistent disturbance. They use the same supervisory activation, protective feedback action, safety-action status, release condition, and shutdown guards described in Section 4.4. The binary system applies $\alpha_k = 1$ whenever $q_k = 1$, whereas the smart system schedules the cycle command α_j^{cmd} while retaining the saturated required intensity in Eq. (20) and the maximum-action requirements described above. Each case is simulated until 50 complete evaluation intervals have been obtained; consequently, the two simulations need not have the same total elapsed time.

As defined by the safety-action activation events in Eq. (51), $\sigma_j = t_{k_j}$ is the activation time associated with the j th transition satisfying $q_{k_{j-1}} = 0$ and $q_{k_j} = 1$. For $\mathcal{E}_j = [\sigma_j, \sigma_{j+1})$, $J_j = \int_{\sigma_j}^{\sigma_{j+1}} \dot{m}_{rel}(t) dt$

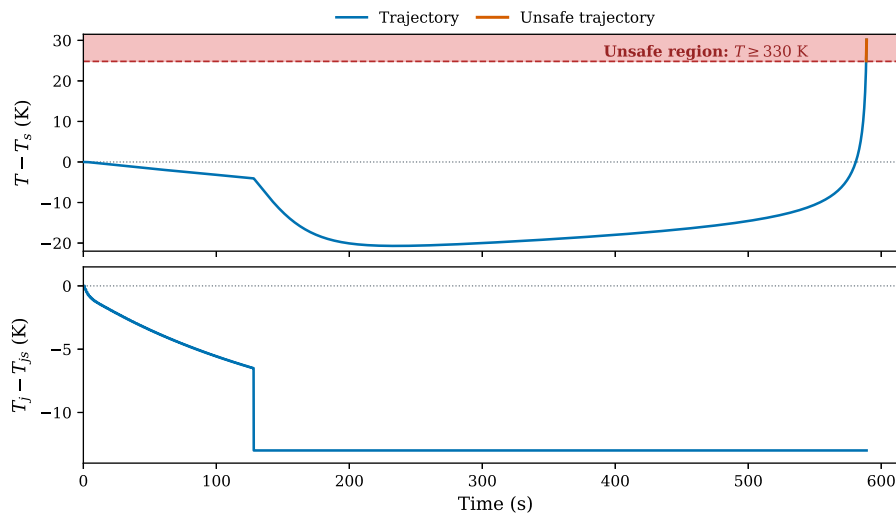


Fig. 5. Reactor and jacket temperature responses corresponding to Fig. 4. Although the jacket temperature reaches its lower bound of 280 K, the reactor temperature reaches a minimum near 284.5 K before crossing $T_{\text{lim}} = 330$ K at approximately $t = 588.7$ s.

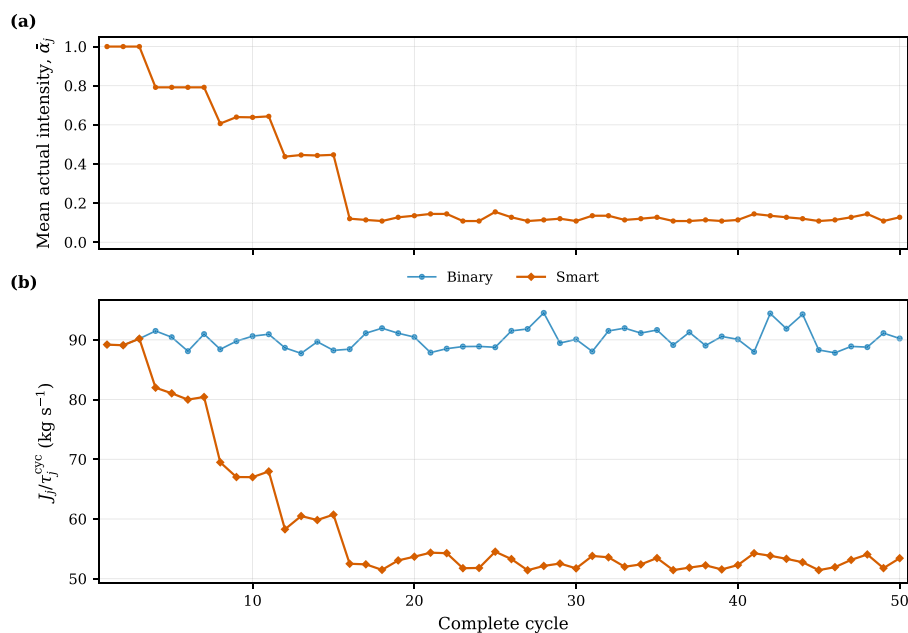


Fig. 6. Cycle-level smart action and material-loss performance under the persistent abnormal disturbance. (a) Mean actual intensity $\bar{\alpha}_j$ during $q_k = 1$ for each complete cycle. (b) Cycle-normalized material-loss metric J_j/τ_j^{cyc} for the binary and smart systems over 50 complete cycles.

denotes the total material discharged during the j th complete cycle, and $\tau_j^{\text{cyc}} = \sigma_{j+1} - \sigma_j$ denotes its duration. The ratio J_j/τ_j^{cyc} is the cycle-normalized material-loss metric used to compare the binary and smart safety systems. The equal injected mass is not included in J_j . Periods during which the relief valve is closed contribute zero to J_j but remain part of τ_j^{cyc} . The same event definition and material accounting are used for both systems.

The smart system yields a final cycle command of $\alpha_{50}^{\text{cmd}} = 0$. The actual sampled intensity nevertheless remains subject to the saturated required intensity and the one-interval maximum capture action.

To distinguish the applied action from its cycle command, define the mean actual intensity during the active portion of the j th complete cycle as $\bar{\alpha}_j = (\sum_{t_k \in \mathcal{E}_j} q_k \alpha_k \Delta) / (\sum_{t_k \in \mathcal{E}_j} q_k \Delta)$. Fig. 6(a) shows $\bar{\alpha}_j$ for each complete cycle. The final-cycle value is $\bar{\alpha}_{50} \approx 0.128$, rather than zero,

because the required safety intensity and the one-interval maximum capture action retain priority over the cycle command. Fig. 6(b) compares the cycle-normalized material-loss metric J_j/τ_j^{cyc} for the two systems. All transition- and override-affected cycles remain included in the physical performance record.

The corresponding concentration and temperature trajectories are shown in Fig. 7. The light curves represent the complete 50-cycle histories, whereas the emphasized curves represent the final complete intervals. Both systems leave Ω_{ρ^*} and undergo repeated excursions under the persistent disturbance. Portions of a trajectory may reenter Ω_{ρ^*} ; however, this does not initiate nominal handoff because the abnormal condition remains unresolved.

The maximum reactor temperatures over the 50 complete intervals are 320.26 K for the binary system and 320.09 K for the smart system. Both remain below the active-safety boundary $T_A = 329.5$ K and the

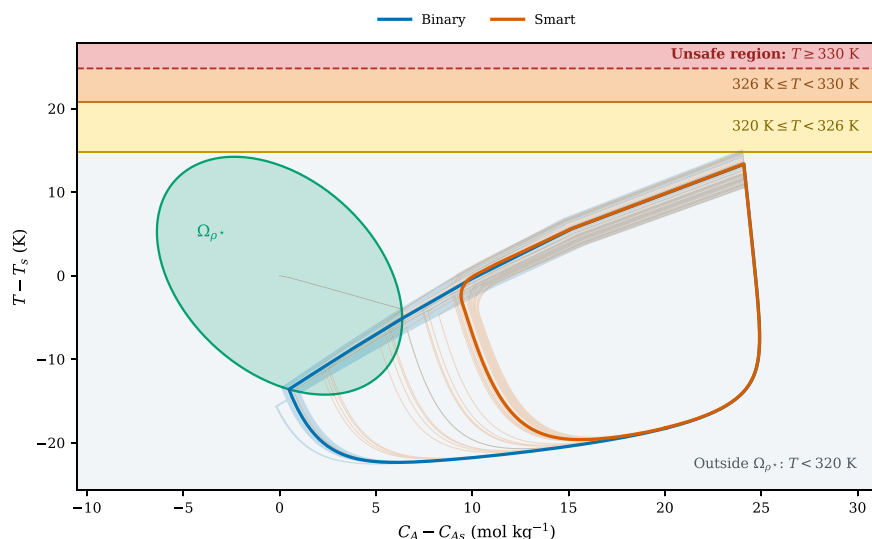


Fig. 7. State-space trajectories of the binary and smart safety systems under the persistent abnormal disturbance. Light curves show the complete 50-cycle histories, whereas the emphasized curves show the final complete intervals. The green region denotes Ω_{ρ^*} , and the temperature-region labels retain their absolute-temperature meanings. Both systems numerically contain the tested trajectories below the physical temperature limit. Entry into Ω_{ρ^*} does not initiate nominal handoff because the abnormal condition remains unresolved. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

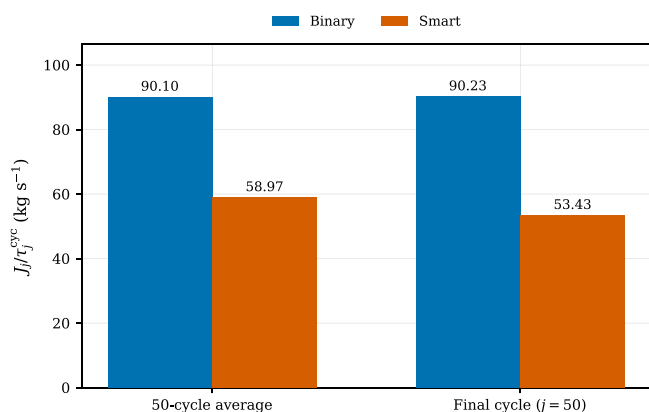


Fig. 8. Comparison of the cycle-normalized material-loss performance. The first pair of bars gives the arithmetic mean of J_j/τ_j^{cyc} over 50 complete cycles, and the second pair gives J_{50}/τ_{50}^{cyc} . Relative to the binary system, the smart system reduces these measures by 34.55% and 40.78%, respectively.

physical limit $T_{lim} = 330$ K. Thus, for this tested disturbance, the smart policy reduces the cycle-normalized material-loss metric without increasing the observed maximum reactor temperature. The horizontal temperature bands in Fig. 7 are labeled using the absolute temperatures 320, 326, and 330 K; they are reference partitions rather than level sets of the history-dependent Safeness Index.

For the aggregate comparison, the 50-cycle arithmetic mean rate is $50^{-1} \sum_{j=1}^{50} J_j/\tau_j^{cyc}$. As shown in Fig. 8, the smart system reduces this mean from 90.10 to 58.97 kg/s, corresponding to a 34.55% reduction. For the final cycle, J_{50}/τ_{50}^{cyc} decreases from 90.23 to 53.43 kg/s, corresponding to a 40.78% reduction. Therefore, the smart safety action substantially reduces the cycle-normalized discharge metric while retaining numerical containment for the tested disturbance.

Finally, the complete evolution from nominal operation through containment, disturbance repair, recovery, and return to nominal operation is evaluated in a 5400 s simulation of the smart safety system. The abnormal inlet concentration is maintained until repair is confirmed at $t_{kf} = 3600$ s, when it is returned to $C_{A0s} = 29.35$ mol/kg. The

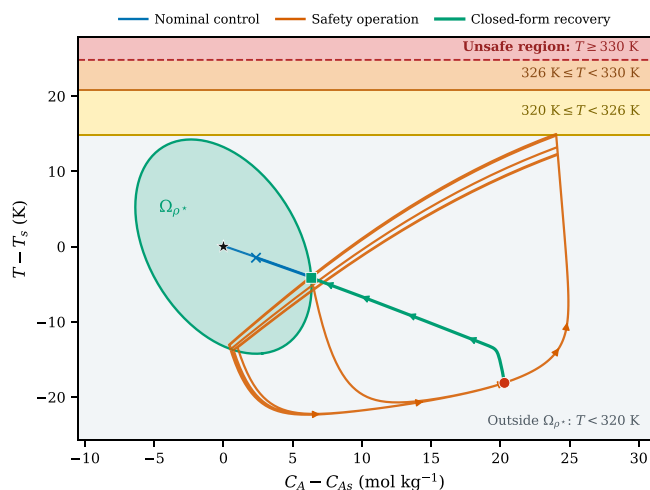


Fig. 9. Complete 0–5400 s state space response of the smart safety system, with repair confirmed at $t_{kf} = 3600$ s and nominal handoff accepted at $t = 3872$ s. Blue trajectory segments correspond to nominal control, the orange curve to safety operation before repair, and the green curve to recovery. The red circle and green square denote the repair and handoff states, respectively; the black star is the nominal equilibrium, and the blue cross is the final state at $t = 5400$ s. The temperature-region labels use absolute temperature and are not Safeness Index level sets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

disturbance is removed completely after this time, but the process state is not reset.

The complete state space response is shown in Fig. 9. Starting from the nominal equilibrium, the state leaves Ω_{ρ^*} , and the supervisor is activated at the next sampling instant, $t = 128$ s. The orange trajectory represents safety operation under the unresolved disturbance. It includes six episodes of the paired relief and injection action, together with intermediate periods during which the safety input is inactive and the protective feedback action acts alone. Although portions of these excursions enter Ω_{ρ^*} , nominal handoff is not permitted before repair has been confirmed. The maximum reactor temperature before repair is 320.15 K, which remains below $T_A = 329.5$ K and $T_{lim} = 330$ K.

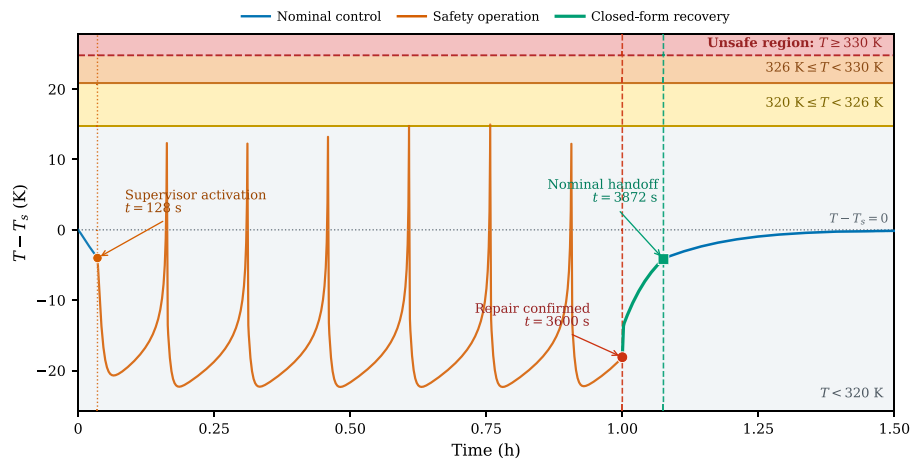


Fig. 10. Temperature-deviation response for the same 0–1.5 h simulation. The ordinate is $T - T_s$, and time is reported in hours. The shaded regions are labeled by their absolute-temperature ranges. Vertical markers identify supervisory activation at 128 s, repair confirmation at 3600 s, and nominal handoff at 3872 s. Blue, orange, and green segments denote nominal control, safety operation, and recovery, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 10 presents the same evolution in time. The ordinate is the temperature deviation $T - T_s$, whereas the shaded regions retain the absolute-temperature meanings shown in the state space figure. The repeated temperature excursions before $t = 3600$ s correspond to containment while the abnormal condition remains unresolved. After repair, the temperature follows the recovery trajectory and then approaches its nominal value after handoff.

At repair confirmation, indicated by the red circle in Fig. 9, the state remains outside Ω_{ρ^*} , with $V(x(t_k)) \approx 7.13 \times 10^4$. Repair confirmation terminates the unresolved-disturbance policy but does not immediately reset the supervisory latch or return the process to nominal operation. Instead, the recovery policy starts from the state reached at repair.

At each recovery sampling instant, the feedback action $u_{R,k}$ is selected first, after which α_k supplies only the additional safety action required by the Lyapunov and temperature conditions. For this trajectory, the temperature condition imposes no positive lower bound on α_k , whereas the positive intensities selected from the Lyapunov condition range from 0.00561 to 0.0184. Consequently, all 272 recovery intervals use the combined feedback and safety action, and no interval uses the feedback action alone. This does not imply that the safety input is required for every admissible recovery trajectory; the same policy selects $\alpha_k = 0$ whenever $u_{R,k}$ alone satisfies both recovery conditions.

The continuous trajectory first enters Ω_{ρ^*} at $t = 3871.97$ s, or 271.97 s after repair. At the next sampling instant, $t = 3872$ s, the state satisfies $V(x) \approx 6998 < \rho^* = 7000$. The Safeness Index condition, SI-LMPC feasibility test, and one-interval handoff checks are also satisfied at the same sampling instant in this simulation. Thus, sampled entry and accepted nominal handoff coincide for this trajectory. This coincidence does not imply that entry into Ω_{ρ^*} is generally sufficient for nominal handoff.

For comparison, Eq. (61) gives a continuous-time entry bound of 566.3 s, and Eq. (62) gives a sampled bound of 567 s, corresponding to 567 sampling intervals. The observed 272 s recovery satisfies both bounds. Following handoff, the supervisory latch is reset, $q_k = 0$, $\alpha_k = 0$, and the nominal SI-LMPC is resumed. No supervisory reactivation or SI-LMPC fallback occurs during the remaining 1528 s. At the end of the simulation, $V(x(5400)) \approx 4.90$ and $T(5400) - T_s \approx -0.14$ K, showing continued motion toward the nominal equilibrium.

The maximum evaluated intersample Lyapunov residual during recovery is $-37.52 < 0$, and no recovery infeasibility, sample-hold violation, actuator chattering, shutdown, or supervisory reactivation is observed. This recovery check uses the saved 0.01 s trajectory; $m_A = 60$ has not been established as a uniform bound over a complete recovery set.

Remark 9. The reduction in the cycle-normalized material-loss metric obtained in this application by testing smaller cycle commands α_j^{cmd} , while enforcing $\alpha_k \geq \alpha_k$, should not be interpreted as a general monotonic relationship between valve opening and cumulative material loss. Although a smaller applied intensity may reduce the instantaneous relief rate, it may require the relief valve to remain open for a longer period and may therefore result in equal or even greater total material loss. Conversely, a larger intensity generally provides a stronger immediate protective effect and may bring the process to a safer condition more rapidly. Hence, the preferred intensity cannot in general be determined from the valve opening alone. The proposed cycle-to-cycle feedback instead adjusts α_j^{cmd} according to completed measurements of J_j/τ_j^{cyc} , while the saturated required intensity $\underline{\alpha}_k$ enforces the prescribed estimated rate condition whenever feasible and the numerical implementation requests shutdown when a required safety condition is infeasible.

5. Conclusion

This work developed an integrated control and safety system architecture for nonlinear safety-critical processes that coordinates feedback regulation and active protection under abnormal disturbances using a safety system with a relief valve, shutdown, and recovery after repair. During nominal operation, the proposed SI-LMPC incorporates safety considerations into the control objective while retaining the Lyapunov condition required for closed-loop stability. Once the process state escapes the closed-loop stability region under a large disturbance that cannot be handled by the control system, the supervisory policy avoids prediction using a potentially unreliable abnormal process model and instead uses sampled measurements together with a certified characterization of the safety input effect to determine the minimum protective action required during each sampling interval to be taken by the safety relief valve. Cycle-to-cycle adjustment can then reduce unnecessary safety system action without violating this certified requirement. Under specific assumptions, the resulting framework establishes robust forward invariance and ultimate boundedness during nominal operation, physical safety throughout certified active safety operation, and finite-time return to the nominal closed-loop stability region after repair through a recovery policy that prioritizes feedback control and supplements it with the safety system input only when necessary. In the MIC hydrolysis CSTR study, feedback cooling alone was unable to contain the tested abnormal disturbance, whereas both the binary and smart safety systems maintained containment. Compared with conventional on/off safety relief valve operation, the proposed

safety system operation policy reduced the reported 50-cycle mean and final-cycle discharge-rate metrics by 34.55% and 40.78%, respectively, without increasing the observed maximum reactor temperature. After repair, the recovery policy returned the process to the nominal closed-loop stability region and completed the nominal handoff 272 s after repair. Overall, the proposed architecture demonstrates that rigorous process protection does not require either prediction with an unreliable abnormal process model or indiscriminate use of full-intensity safety action, thereby providing a systematic means of reconciling physical safety, reduced material loss, and recovery within a unified supervisory framework.

CRedit authorship contribution statement

Xiaodong Cui: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Arthur Khodaverdian:** Writing – original draft, Methodology, Conceptualization. **Panagiotis D. Christofides:** Writing – review & editing, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The corresponding author is an Editorial Board Member for this journal and was not involved in the editorial review or the decision to publish this article.

Acknowledgments

Financial support from the National Science Foundation, CBET-2227241, is gratefully acknowledged. PDC would like to acknowledge Professor Arkun's numerous contributions to process control, and in particular, his early work on feedforward/feedback control of nonlinear processes with disturbances that provided an inspiration to consider various disturbance types in the present work. Beyond his technical contributions, Professor Arkun set an impeccable example of leadership and collegiality for over five decades in the field of process control that greatly fostered the growth of the field and the success of generations of process control researchers.

References

- Ahooyi, T.M., Soroush, M., Arbogast, J.E., Seider, W.D., Oktem, U.G., 2016. Model-predictive safety system for proactive detection of operation hazards. *AIChE J.* 62, 2024–2042.
- Albalawi, F., Durand, H., Christofides, P.D., 2017. Process operational safety using model predictive control based on a process Safeness Index. *Comput. Chem. Eng.* 104, 76–88.
- Albalawi, F., Durand, H., Christofides, P.D., 2018. Process operational safety via model predictive control: Recent results and future research directions. *Comput. Chem. Eng.* 114, 171–190.
- Ball, R., 2011. Oscillatory thermal instability and the Bhopal disaster. *Process Saf. Environ. Prot.* 89, 317–322.
- Cheng, R., Orosz, G., Murray, R.M., Burdick, J.W., 2019. End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33, pp. 3387–3395.
- Christofides, P.D., Scattolini, R., Muñoz de la Peña, D., Liu, J., 2013. Distributed model predictive control: A tutorial review and future research directions. *Comput. Chem. Eng.* 51, 21–41.
- Cui, X., Khodaverdian, A., Christofides, P.D., 2026a. Control Lyapunov-barrier function-based reinforcement learning of nonlinear processes. *Comput. Chem. Eng.* 213, 109777.
- Cui, X., Khodaverdian, A., Christofides, P.D., 2026b. Reinforcement learning-based control of nonlinear systems: Economic optimization and operational safety. *AIChE J.* 72, e70479.
- Cui, X., Peters, D., Wang, Y., Çıtmacı, B., Richard, D., Morales-Guio, C.G., Christofides, P.D., 2024. Modeling and control of a protonic membrane steam methane reformer. *Chem. Eng. Res. Des.* 212, 493–519.
- Daş, E., Murray, R.M., 2022. Robust safe control synthesis with disturbance observer-based control barrier functions. In: *2022 IEEE 61st Conference on Decision and Control*. CDC, pp. 5566–5573.
- Dhara, V.R., Dhara, R., 2002. The Union Carbide disaster in Bhopal: A review of health effects. *Arch. Environ. Health: Int. J.* 57, 391–404.
- Easteal, A.J., Back, P.J., Woolf, L.A., 1997. PVT property measurements for liquid chlorobenzene and 1,2-dichlorobenzene from (278 to 338) K and (0.1 to 300) MPa. *J. Chem. Eng. Data* 42, 1261–1265.
- Ellis, M., Durand, H., Christofides, P.D., 2014. A tutorial review of economic model predictive control methods. *J. Process Control* 24, 1156–1178.
- Emelianova, A., Gor, G.Y., 2021. Molecular simulations of vapor–liquid equilibrium of isocyanates. *J. Phys. Chem. B* 125, 12528–12538.
- Hace, I., 2013. The pressure relief system design for industrial reactors. *J. Ind. Eng.* 2013, 453509.
- Hristov, H.V., Mann, R., 2002. Fluid mixing and the safe quenching of a runaway reaction in a stirred autoclave. *Chem. Eng. Res. Des.* 80, 872–879.
- Kanavalau, A., Lall, S., 2022. Thermal runaway avoidance using Hamilton–Jacobi reachability and model predictive control. *Comput. Chem. Eng.* 157, 107605.
- Kummer, A., Varga, T., Nagy, L., 2020. Semi-batch reactor control with NMPC avoiding thermal runaway. *Comput. Chem. Eng.* 134, 106694.
- Leung, J.C., 1986. Simplified vent sizing equations for emergency relief requirements in reactors and storage vessels. *AIChE J.* 32, 1622–1634.
- Leveson, N.G., Stephanopoulos, G., 2014. A system-theoretic, control-inspired view and approach to process safety. *AIChE J.* 60, 2–14.
- Lin, Y., Sontag, E.D., 1991. A universal formula for stabilization with bounded controls. *Syst. Control Lett.* 16, 393–397.
- Occupational Safety and Health Administration, 1992. 29 CFR 1910.119: Process Safety Management of Highly Hazardous Chemicals. United States Department of Labor.
- Oruganti, P.S., Naghizadeh, P., Ahmed, Q., 2024. Robust control barrier functions for sampled-data systems. *IEEE Control Syst. Lett.* 8, 103–108.
- Peters, D., Cui, X., Wang, Y., Morales-Guio, C.G., Christofides, P.D., 2026. Model predictive control of an experimental protonic membrane steam methane reforming system. *AIChE J.* 72, e70105.
- Qin, S.J., Badgwell, T.A., 2003. A survey of industrial model predictive control technology. *Control Eng. Pract.* 11, 733–764.
- Senthilvelan, K., Shaik, S.M., Wong, W.C., Sugiman, D., 2025. From theory to practice: Implementing safe operating limits in major hazard facilities. *Process Saf. Prog.* 44, 553–563.
- Stauffer, T., Chastain-Knight, D., 2021. Do not let your safe operating limits leave you S-O-L (out of luck). *Process Saf. Prog.* 40, e12163.
- Toro, J.C.O., Dobrosz-Gómez, I., García, M.A.G., 2016. Dynamic modeling and bifurcation analysis for the methyl isocyanate hydrolysis reaction. *J. Loss Prev. Process Ind.* 39, 106–111.
- Torré, J.-P., Fletcher, D.F., Lasuye, T., Xuereb, C., 2008. An experimental and CFD study of liquid jet injection into a partially baffled mixing vessel: A contribution to process safety by improving the quenching of runaway reactions. *Chem. Eng. Sci.* 63, 924–942.
- U.S. Chemical Safety and Hazard Investigation Board, 2023a. Fatal Chemical Decomposition Reaction and Explosion at Optima Belle LLC. Investigation Report 2021-02-I-WV, U.S. Chemical Safety and Hazard Investigation Board, Washington, DC.
- U.S. Chemical Safety and Hazard Investigation Board, 2023b. Storage Tank Fire at Intercontinental Terminals Company, LLC (ITC) Terminal. Investigation Report 2019-01-I-TX, U.S. Chemical Safety and Hazard Investigation Board, Washington, DC.
- U.S. Chemical Safety and Hazard Investigation Board, 2025. Marathon Martinez Renewable Fuels Fire. Technical Report, U.S. Chemical Safety and Hazard Investigation Board.
- U.S. Chemical Safety and Hazard Investigation Board, 2026a. Final Report on Fatal 2024 Explosion at Givaudan Facility in Louisville, Kentucky. Technical Report, U.S. Chemical Safety and Hazard Investigation Board.
- U.S. Chemical Safety and Hazard Investigation Board, 2026b. Investigation Report on the 2023 Explosion and Toxic Ethylene Oxide Release at Dow Plant in Plaquemine, Louisiana. Technical Report, U.S. Chemical Safety and Hazard Investigation Board.
- Wang, Y., Ji, Z., Cao, Y., Yang, S.-H., 2023. Safety critical variable analysis for process systems. *Ind. Eng. Chem. Res.* 62, 21704–21720.
- Wang, Y., Wu, Z., 2024. Control Lyapunov-barrier function-based safe reinforcement learning for nonlinear optimal control. *AIChE J.* 70, e18306.
- Wu, Z., Albalawi, F., Zhang, Z., Zhang, J., Durand, H., Christofides, P.D., 2019. Control Lyapunov-Barrier function-based model predictive control of nonlinear systems. *Automatica* 109, 108508.
- Zhang, Z., Wu, Z., Durand, H., Albalawi, F., Christofides, P.D., 2018a. On integration of feedback control and safety systems: Analyzing two chemical process applications. *Chem. Eng. Res. Des.* 132, 616–626.
- Zhang, Z., Wu, Z., Durand, H., Albalawi, F., Christofides, P.D., 2018b. On integration of model predictive control with safety system: Preventing thermal runaway. In: *Computer Aided Chemical Engineering*, vol. 44, pp. 2011–2016.
- Zhang, Z., Wu, Z., Rincón, D., Christofides, P.D., 2019. Integrating safeness index-based model predictive control and safety relief valve activation for operational safety of chemical processes. In: *Proceedings of American Control Conference*. Philadelphia, Pennsylvania, pp. 995–1001.